

NIH at the crossroads

The director of the US National Institutes of Health has laid out a plan that would align the world's largest biomedical research agency more closely to the future shape of the life sciences. It deserves political support.

The National Institutes of Health (NIH) has reached a critical moment in its history. After a remarkable surge in its budget — from some \$13 billion in 1998 to \$27 billion this year — the agency will be under more pressure than ever before to deliver results. With expenditure on this scale, the time when the NIH can operate more or less independently from its parent Department of Health and Human Services, or even from the White House, may be drawing to a close, raising new questions about the agency's scientific integrity and mission. After years of enthusiastic support in the US Congress, which sets the agency's budget and has shielded it from political interference, the NIH can look forward to a period when it will get more scrutiny — and less largesse — on Capitol Hill.

That's the situation facing NIH director Elias Zerhouni, the former radiologist and administrator at Johns Hopkins University who took over the directorship of the NIH in May last year. Biologists inside and outside the NIH may feel some regret that Zerhouni lacks the overpowering scientific resumé of his predecessor, Harold Varmus. But right now, outstanding administrative flair may be what the expanded agency needs of its director.

Researchers are looking to Zerhouni to provide just that. On 30 September he reached the most significant landmark so far in his directorship, releasing the NIH Roadmap for Medical Research, which, despite a thorough consultation process, bears his clear personal imprint. The plan is shaped to help the NIH do the kind of science — particularly 'translational research', from the laboratory to the bedside — that Zerhouni considers to be most critical to the agency's central mission of improving US public health.

The plan includes initiatives to foster the development of new medical technologies, including some that could aid the high-throughput identification of protein structures. It would create databases and centres, such as a chemical library and institutes for nanomedicine and biomedical computing. The roadmap also instigates a major push to back more interdisciplinary research, and train more people to do it. And it emphasizes speedier and more all-encompassing clinical research, for example by creating regional centres for translational research and better integrating ordinary physicians and their patients into networks of clinical trials.

Backing groups

The most obvious potential drawback of this ambitious plan is that it may end up diverting money from the individual-investigator grants that have always been the mainstay of the NIH, providing support instead for larger, perhaps more permanent, teams doing interdisciplinary work. Zerhouni says that this merely reflects modern reality, and that the new facilities and structures to be created will make it easier for individual principal investigators to do their jobs. However, efforts to back large groups of researchers and semi-permanent centres of excellence have, in many countries over many decades, shown themselves to be less efficient, less merit-focused and more subject to political interference in the allocation of funds, than competitive, peer-reviewed grants for individual researchers. But even if the plan is implemented in full, individual-investigator

grants will continue to dominate the NIH's portfolio. The roadmap's contents seem to be a reasonable reflection of the changing currents in biomedical research, which call for systematic and interdisciplinary approaches that are not adequately addressed under the NIH's existing structure.

Another aspect of the plan that has stirred discussion is the creation of ten Director's Innovator Awards, which will give \$500,000 in funds to individual scientists selected by a peer-review panel. Zerhouni says that this scheme is intended to encourage risky, path-breaking science. But in the absence of clarity over just how this selection will work, there must be some concern that such awards will go to investigators who are already well established.

Ties with industry

Critics are also questioning aspects of proposals to make it easier for industry to work with the NIH. The roadmap will create a Director's Liaison for Public-Private Partnerships to act as a single point of contact for all of the agency's work with the private sector. But if, as Zerhouni seems to envisage, the agency develops even closer ties with the pharmaceutical industry, questions should be asked in the Congress, where there are some supporters of greater collaboration, about what the pharmaceutical companies are doing in return. For example, might it cap the premium prices that Americans pay for medicines that have often been developed on the back of NIH-funded research projects?

Similarly, the roadmap's vision of a simpler and more transparent process for the regulation of clinical trials may collide with public demands to push regulation in the opposite direction, after recent cases in which patients have died during clinical trials under questionable circumstances.

The roadmap was developed in consultation with scientists both within and outside the NIH. Zerhouni has talked since he arrived at the NIH about the need to foster innovative technologies, to encourage more team science, and to streamline clinical research — but he seems to have made an honest effort to incorporate the views of the community. And there is no sign that Zerhouni has allowed either the health department or the White House to impose their political priorities on the plan.

Zerhouni has won public support for this vision from the fiercely independent directors of the agency's 27 separate institutes and centres, who have promised to chip in money from their own budgets to fund interagency efforts that the roadmap says will cost \$125 million next year, building up to \$2 billion by 2009. It remains to be seen, of course, what these promises will be worth if money is tight at the institutes when the time comes.

At a congressional hearing on 2 October, Zerhouni fielded a few questions about the roadmap. But he faced just as many, if not more, from lawmakers who were concerned that the NIH had funded research on prostitutes, on transgendered Native Americans and on the sexual habits of older men. Political skills on Zerhouni's part will clearly be as essential for the NIH's own good health as they were for his predecessors. Meanwhile, he deserves credit for laying out a clear vision of what the future of his agency should be. ■





Noise annoys
Moves to relax sonar rules spark marine protest
p549



Cabbing fever
Ig Nobel prizes give study of taxi-drivers the thumbs up
p550



Judgement day
Everglades project flounders after judge is ousted
p551



Weather report
Japan throws the switch for high-res climate model
p552

Magnetic pioneers net Nobel for putting medicine in the picture

Helen Pearson, London

For allowing doctors to peer into previously impenetrable parts of the human body, two physicists have scooped the 2003 Nobel Prize in Physiology or Medicine. Paul Lauterbur of the University of Illinois in Urbana and Peter Mansfield of the University of Nottingham, UK, share the prize for their contribution to the development of magnetic resonance imaging (MRI).

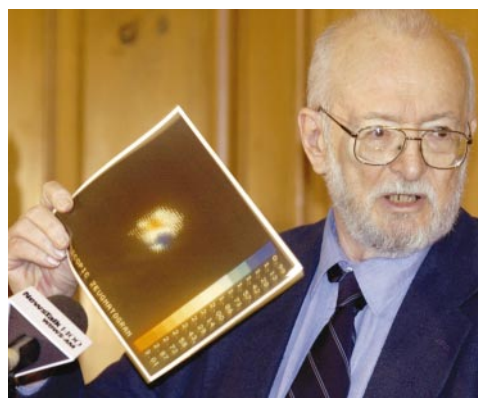
Before the advent of MRI in the 1970s, doctors struggled to examine organs and soft tissues in people using X-rays. Today, MRI clinicians and medical researchers alike can readily view details of the internal structure of patients' entire bodies.

Lauterbur and Mansfield kick-started this revolution in medical imaging by tweaking a technique called nuclear magnetic resonance (NMR), primarily used for investigating molecular structures. When placed in a magnetic field, the nuclei of certain atoms, including hydrogen, line up. Subjected to a blast of radio waves, they absorb this radiation and emit it again as an NMR signal — from which the structure of the molecules in which the atoms sit can be deciphered.

While working at the State University of New York at Stony Brook, Lauterbur came up with the idea of using a graded magnetic field that increased in strength from one side of an object to the other to create a two-dimensional image of the object's internal structure. This is based on the idea that nuclei absorb and emit radio waves at different frequencies, depending on the strength of the magnetic field in which they are held.

In a seminal paper, Lauterbur produced crude images of two glass capillaries filled with water (P. C. Lauterbur *Nature* **242**, 190–191; 1973). His manuscript was initially rejected by *Nature*, but his protests proved successful. "I congratulate you on eventually giving in to my anguish," he told *Nature* this week, hours after the prize was announced.

From the beginning, Lauterbur believed



Winners: Paul Lauterbur (left) with the first magnetic resonance image and Peter Mansfield by the machine it inspired.



that the technique could be used to image live animals, rather than killing them to remove and study their tissues. "I thought there must be another way," he says.

But at the time, it was difficult to envisage the technique evolving to give us the sophisticated medical images that are now routine — nearly 60 million of them were taken in 2002. "No one could imagine doing this to people," explains radiologist Paul Bottomley of Johns Hopkins University in Baltimore, Maryland.

It was Mansfield who, in a series of papers, provided imaginative leaps to turn MRI into a useful clinical tool. "His talent is tenacity and coming up with good ideas," says Bottomley.

The first advance made by Mansfield was to show that it is possible to selectively image a particular two-dimensional slice of an object held in a graded magnetic field by delivering the blast of radio waves at a particular frequency. Second, he developed techniques to manipulate the applied magnetic fields and algorithms to interpret the resulting cacophony of radio signals that enabled images to be pieced together in just a matter of seconds.

Mansfield's work has allowed researchers

and clinicians to take snapshots of fast-moving events, such as the beating of the heart. It has also been vital in the development of brain scanning by functional MRI, in which mental activity can be tracked by monitoring changes in blood flow. "It's been critical," says Paul Matthews, director of the

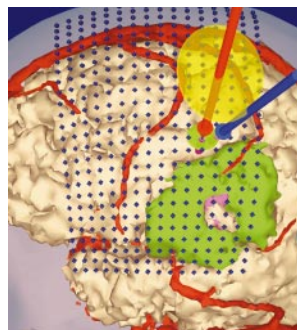
Centre for Functional MRI of the Brain at the University of Oxford, UK.

A Nobel for MRI has been long anticipated but, because its development has involved many steps, deciding who to reward won't have been easy. "It's been a hot potato for the Nobel prize committee," suggests John Griffiths of St George's Hospital Medical School, London, who uses MRI to diagnose and study cancer — one of its most

important clinical applications.

In the end, the Nobel committee seems to have awarded the prize specifically for the initial developments involving the use of graded magnetic fields. But experts say that this does not preclude recognition of other contributors to MRI in the future. "It leaves the way open," says Stephen Keevil, a magnetic resonance physicist at Guy's Hospital, London.

♦ www.nobel.se



MRI in action: a brain tumour (green) is marked for surgery.

S. TAN/AP

MIT ALAB/SURGICAL PLANNING LAB/BRIGHAM & WOMEN'S HOSP/SPL

Super-cool theories secure physics prize

Sarah Tomlin, London

Three physicists who have wrestled with the explanations behind intriguing quantum phenomena share this year's Nobel Prize in Physics.

Two Russians, Alexei Abrikosov of the Argonne National Laboratory in Illinois and Vitaly Ginzburg, the retired head of the theory group at the P. N. Lebedev Physical Institute in Moscow, are rewarded for their theoretical explanation of a form of superconductivity. Anthony Leggett of the University of Illinois at Urbana-Champaign is recognized for his work on a type of superfluid.

A superfluid forms when a fluid, such as liquid helium, effectively loses all of its viscosity and flows without any resistance. If that fluid is made up of electrons rather than atoms, the effect — electrical conductivity without resistance — is called superconductivity. Both phenomena depend upon all of the fluid's constituent atoms or 'free' electrons dropping into the same quantum state — something that happens only at low temperatures.

Superfluids are divided into more than one type of material and behaviour. They were discovered in the 1930s, when helium-4 was cooled to very low temperatures and began to flow without resistance. It was later



Members of the Nobel committee on physics announce this year's prize recipients.

discovered, almost by accident, that helium-3 could also become a superfluid, although at a much lower temperature — about 1,000 times lower than that needed for helium-4. This was difficult to explain, because helium-3 atoms belong to a class of particles known as fermions, which don't like to share the same quantum state. It was Leggett who, in the 1970s, succeeded in explaining how this

happens: the atoms first form into pairs, which then act like single particles that can occupy the same quantum state.

Superconductivity had long been known also to depend on a similar pairing of electrons. But until Abrikosov and Ginzburg's theory, building on work by a previous Nobel winner, Lev Landau, theorists struggled to explain the behaviour of a class of superconductors, called 'type II', which can remain superconducting in the presence of a magnetic field.

Their theory is now textbook material. It is particularly important for understanding the properties of superconducting magnets, such as those used in particle accelerators. "People use the Ginzburg-Landau theory all the time," says Gil Lonzarich, a physicist at the University of Cambridge, UK. "I'm looking at my board right now, and there is a Ginzburg-Landau equation on it."

Although the three theorists share similar interests, says Lonzarich, they differ in their approach to research. Leggett is "an intuitive thinker", says Lonzarich, whereas Abrikosov is a more formal mathematician. He describes Ginzburg as one of those rare people who can do both formal and more intuitive thinking.

The three theorists helped to change the field of condensed-matter physics by recognizing the importance of interactions between atoms or electrons — the behaviour of strongly interacting particles together can be more important than individual particle behaviour at the quantum level. This "led to major changes in thinking", says Lonzarich, who argues that the recognition of the work by Abrikosov and Ginzburg, done in the 1950s, is particularly long overdue. ■

The Nobel chemistry prize was announced after *Nature* went to press. For coverage see www.nature.com/nature

Biologists join physics preprint club

Declan Butler

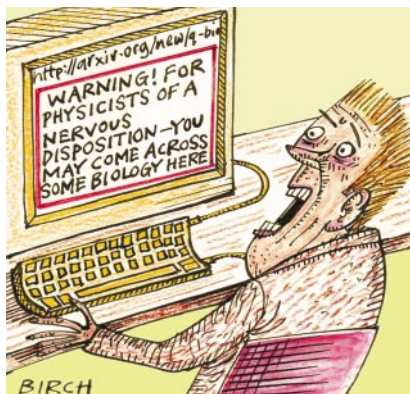
The ArXiv preprint server — physicists' favourite place for early circulation of their results — has branched out into biology.

Last month, the server's managers created q-bio, an archive for quantitative biology. The move reflects the fact that ArXiv's traditional constituency of physicists, mathematicians and computer scientists is increasingly working on biological problems, says its founder Paul Ginsparg of Cornell University in New York state.

The proportion of biology papers on ArXiv has been growing steadily, reaching 8% of the 4,500 submissions received last year. But until now, these papers have been scattered across various subdisciplines on ArXiv. The new arrangement will regroup existing content, and provide a dedicated area for quantitative biology.

Ginsparg says he hopes that the move will help to "nucleate something for newcomers" from biological disciplines. The big test, he says, will be whether biologists will follow physicists in being comfortable with circulating their 'big' papers on the archive prior to formal publication.

But papers on ArXiv are not peer-

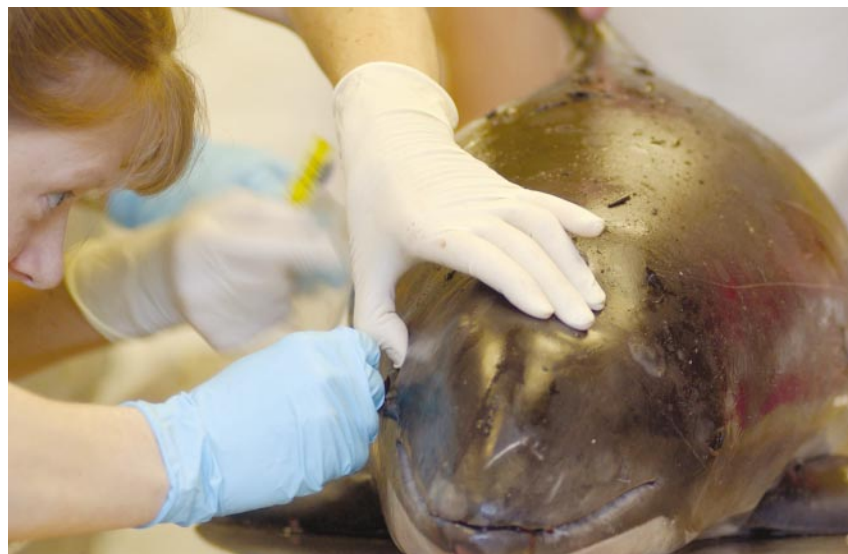


reviewed, and there is concern this could create problems if medical papers are accessed by physicians or patients. Ginsparg says he is not really worried by this, adding that the biology papers on the archive are "generally as uncontroversial as for physics".

One of the few existing biomedical preprint servers — the BMJ's Netprints server — features a 'health warning' advising 'casual readers' not to act on its contents. ArXiv has no plans for such a warning, although an advisory panel does screen submissions. ■
♦ arxiv.org

Scientists split over regulations on sonar use

R. WURZER/AP



Dead in the water: are navy sonar operations to blame for porpoise deaths off Washington state?

Rex Dalton, San Diego

The US Congress is considering proposals that will make it easier to get permission to use high-volume sonars in the ocean — just as fresh evidence suggests that their noise can harm marine mammals.

Capitol Hill is looking at two measures to loosen the 1972 Marine Mammal Protection Act (MMPA), which sets guidelines for noisy experiments in the oceans. One would simplify the rules, making it easier to get permission to do the experiments. The second would exempt the US Navy from the regulations on the grounds of national security.

The changes are supported by the navy and by some geophysicists, who want to use noise-generating devices to study geological formations on the ocean floor. But they are strongly opposed by many marine biologists. “There is a huge split over the issue,” says

John Hildebrand, who studies marine mammal acoustics at the Scripps Institution of Oceanography in La Jolla, California.

In a Brief Communication in this issue of *Nature*, a team led by Paul Jepson of the Institute of Zoology in London concludes that 14 whale deaths off the Canary Islands last year may have been caused by decompression sickness after the animals shot to the surface to escape sonars during Spanish-led international naval exercises (page 575). The team says the sonar appears to have caused gas bubbles to form in the blood, damaging the whales’ livers and kidneys.

Experts say that the study provides some of the most direct evidence to date that sonars can kill marine mammals. “This report has the potential to be the ‘smoking gun’ on the cause of sonar-related mammal strandings,” says Hildebrand. “It certainly

focuses on the potential dangers of sonar, which need to be thoroughly investigated.”

On 25 September, however, the oceans subcommittee of the House of Representatives Committee on Resources passed a ‘reauthorization’ of the MMPA that would give government agencies more freedom to permit experiments in the oceans. Agencies could also ask for stronger proof that a study might cause damage. The bill will soon be considered by the full committee, where some Democrats will seek to tighten its provisions.

Meanwhile, a House–Senate conference committee is due to consider a 2004 military spending authorization bill that would exempt the navy from the sonar rules. “We fear something bad” is going to come out of the conference, says Karen Wayland, a geologist with the National Resources Defense Council, an environmental group that took successful court action to block the navy’s use of some sonar devices (see *Nature* 425, 6; 2003).

John Orcutt, a geophysicist and deputy director of Scripps, says that he favours the proposed modifications to the MMPA, so that researchers can secure permits more easily than at present. “The process is tremendously flawed,” he says. Orcutt is worried, however, that the navy exemption would encourage the military to do all of its own experiments and stop supporting external researchers, who would remain constrained by the law.

But marine biologist Ken Balcomb, director of the Center for Whale Research in Friday Harbor, Washington, says recent deaths and strandings of marine mammals should persuade physical oceanographers — and Congress — to worry more about protecting these animals instead of loosening research regulations. ■

European Commission dips toe into military research

Marika Willerroider, Munich

Plans for a European defence research agency took a tentative step forward this week, when the European Commission (EC) announced that it would fund E65 million (US\$76 million) of security-related research and development over three years.

The EC will spend the money on research into what it terms “non-offensive” military capabilities, such as non-lethal weapons for riot control or early-warning systems against bioterrorism. E9 million will be released to kick-start research in 2004. The rest will need approval by the European parliament.

The pilot programme — the EC’s first

foray into military research — will fund joint academic–industry projects, and a call for proposals will go out in January, an EC spokesman said. The European Security Research Programme will be jointly administered by the EC directorates for research and enterprise.

But the project could form the nucleus of a much wider involvement in military research. On 6 October, a group of advocates including Lord Robertson, the NATO secretary general, presented a plan for a fully-fledged European Union (EU) defence research agency to EC officials. The meeting also included the chief executives of

electronics companies Siemens and Ericsson, and the European Aeronautic Defence and Space Company. Europe’s military-equipment industry has been lobbying for years for such an outfit, possibly along similar lines to the US Defense Advanced Research Projects Agency (see *Nature* 421, 465; 2003).

Advocates would like to see it up and running by 2007, but this is unlikely as the new agency will require approval from the European parliament, says Elly Plooi-j-van Gorsel, a Dutch member of the parliament. She adds that military research needs funds beyond the existing EU research budget. ■

Spoof Nobels take researchers for a ride

Steve Nadis, Boston

Eleanor Maguire was in an apprehensive mood. She had just flown into Boston, Massachusetts, to collect an Ig Nobel Prize in Medicine at Harvard University. But just before the ceremony, she was having second thoughts. "What have I gotten myself into?" she asked herself aloud.

Earlier this year, Maguire, a neuroscientist at University College London, had barely heard of the Ig Nobels, which were awarded for the 13th time on 2 October by a self-appointed panel for "research that cannot or should not be reproduced". Maguire had no idea why she'd won — although the title of a paper she co-authored in 2000, "Navigation-related structural change in the hippocampi of taxi drivers" (*Proc. Natl Acad. Sci. USA* **97**, 4398–4403), might have offered her a clue.

In the paper, Maguire and her colleagues showed that the posterior hippocampus, thought to be a navigation centre in the brain, is enlarged in London taxi-drivers. What's more, the volume of the hippocampus correlates with years of experience in the cab. This research was no joke, as it suggested a capacity for structural plastic changes in healthy human brains — a finding with important clinical implications. Yet it plunged Maguire into an annual prize-giving ceremony best-known for its frivolity and sophomoric gags.

This year, one lucky spectator won a date with a proper Nobel laureate, Richard Roberts (Physiology or Medicine, 1993), also known as Mr December in the 1997 "Studmuffs of Science" calendar. Another got to meet Stephen Wolfram, who promised to explain his intense 1,200-page tome *A New Kind of Science* over a cup of tea.



Fare play, guv: London cabbies have a big hippocampus, according to an Ig Nobel-winning study.

In a "24/7 nanolecture", geneticist Eric Lander summarized the Human Genome Project in 24 seconds — "The genome cost US\$3 billion and gave us three billion letters: one dollar a letter. Such a deal!" — and then, rather more successfully, in seven words: "Genome. Bought the book. Hard to read."

The physics Ig Nobel went to Australian investigators for "An analysis of the forces required to drag sheep over various surfaces" (*Appl. Ergon.* **33**, 523–531; 2002). Yukio Hirose of Kanazawa University in Japan captured the chemistry award for his studies of a bronze statue that failed to attract pigeons.

Researchers at Stockholm University, Sweden, claimed the interdisciplinary prize

for "Chickens prefer beautiful humans" (*Hum. Nature* **13**, 383–389; 2002). And C. W. Moeliker of the Natuurmuseum in Rotterdam, the Netherlands, won the biology prize for documenting the first known case of homosexual necrophilia in the mallard duck (*Deinsea* **8**, 243–247; 2001).

In her acceptance speech, Maguire told the crowd that she got a 50% discount on a cab fare "for showing that taxi-cab drivers are special" — perhaps the first instance of anyone benefiting materially from winning an Ig. At the ceremony's end, she said: "Some of the other winners did amazing things with sheep, pigeons, chickens and mallard ducks. But for now I'll stick with taxi-drivers." ■

Geologists seek to put an end to blind dates

Tom Clarke

Earth scientists have decided it is time to talk time. At a meeting in Washington DC last week, experts in mass extinctions, ancient climate and the art of dating rocks got together to work out plans for a more accurate and complete geological timescale.

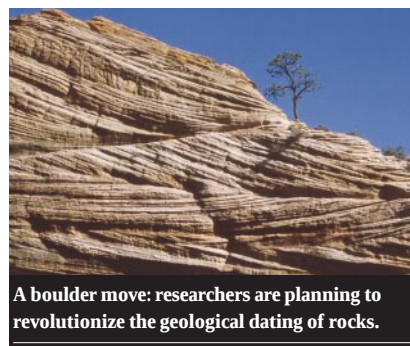
The researchers plan to establish an international network of laboratories that would use agreed standard procedures for dating rocks, and to which all Earth scientists would send samples of agreed quality from important sites. Over time, that would establish a database of reproducibly accurate dates for everything recorded in the Earth's rocks, from the evolution of ancient life to rapid climate-change events.

The project could take 15 years, and could fill in many gaps in our knowledge of

Earth's history, says Samuel Bowring, a geochronologist from the Massachusetts Institute of Technology in Cambridge and one of the meeting's organizers. "The record is hopelessly incomplete," he says.

That is partly because the most accurate techniques for dating are work-intensive and require more skill and money than most labs can spare. So researchers often simply estimate rock ages by comparing the fossils found in one stripe of rock to another of known age.

What's more, many historically important layers of rock were dated before the more accurate techniques were invented. Until recently, for example, the Cambrian explosion, from which most animal species emerged, was thought to have occurred some 575 million years ago over an unknown



A boulder move: researchers are planning to revolutionize the geological dating of rocks.

period of time. But the use of highly accurate uranium-lead isotope measurements allowed Bowring's lab to show that it actually began 544 million years ago and lasted just a few million years (*Science* **261**, 1293–1298; 1993). Such information is crucial for

Judge's sacking rocks Everglades clean-up

Betsy Mason, Washington

The Florida Everglades clean-up effort — one of the world's largest and most hotly debated environmental projects — is in trouble this week, after the judge charged with overseeing the scheme was kicked out for alleged political bias.

The ousting on 23 September of Judge William Hoeveler of the US District court at Fort Lauderdale is the culmination of six months of bitter wrangling over desirable phosphorous levels in the Everglades. The 5,000-square-kilometre wetland in southern Florida is home to at least 60 endangered species, including the American crocodile.

The removal of Hoeveler — largely brought about by sugar growers whose farms north of the Everglades are the source of phosphorus flowing into the swamp — has alarmed some supporters of the \$8-billion clean-up effort. "The scientific system has been undercut," claims ecologist Curt Richardson of Duke University in Durham, North Carolina.

In its natural state, the Everglades supported a unique ecosystem made up of plants and animals that thrive in nutrient-poor conditions. Adjusting the nutrient balance can cause radical changes to the ecosystem, including the invasion of cattails, which overrun native saw-grass and choke the flow of water through the region.

Ecologists want phosphorus levels to be kept below 10 parts per billion (p.p.b.), but levels are as high as 80 p.p.b. in some water flowing into the Everglades. As a result, a 'phosphorous front' has been gradually encroaching on the wetlands at around a kilometre per decade, says ecologist Evelyn Gaiser of Florida International University in



Bittersweet business: a flow of phosphorus from sugar farms poses a threat to Florida's wetlands.

Miami. "Every day counts," says Dan Childers, who studies the Everglades ecosystem at Florida International University in Miami. "As we wait, more and more of the Everglades will be impacted."

When the restoration project began in 1992, Florida's state government agreed to bring the levels of phosphorous in water entering the Everglades down to 10 p.p.b. by 2006.

Since then, sugar farmers — who avoid using phosphorus-based fertilizers, but whose water-management practices cause the release of the chemical — have greatly reduced the amount of phosphorus leaving their farms. But the 2006 deadline is looming, and most experts agree that it is unlikely

to be met. So in May, a law supported by the farmers to extend the deadline until 2016 was passed by the Florida legislature and signed by state governor Jeb Bush, brother of US president George W. Bush.

Hoeveler — whose court has overseen the restoration project from the outset — was quoted in the Florida press as saying that he was "deeply troubled by the content of the bill", and that Governor Bush had allowed the project to "fall into the hands of those who don't like the Everglades". Under a federal code of ethics that prohibits judges from appearing biased or commenting publicly on the merits of a pending case, the US Sugar Corporation of Clewiston, Florida, successfully filed for the judge's removal.

"The principle behind the original settlement was that no contribution to pollution was acceptable," says Neal McAliley, a former US Justice Department attorney on the Everglades case. "Implicit in the new law is the idea that some pollution is acceptable. In my view that represents a shift in thinking at a policy level."

The changes leave ecologists worried that delays could now undermine the entire restoration. But Gary Goforth, an environmental engineer at the South Florida Water Management District, the state agency that is implementing the project, says that most aspects of it are still ahead of schedule.

"It's not a date that matters," says Judy Sanchez, director of corporate communications at US Sugar. "It's progress that matters. And progress has been extraordinary."

Environmentalists remain unconvinced, however. "The sugar industry's interest is and always has been delay," says Thom Rumberger, chair of the Everglades Trust. "This is the ultimate delay."

palaeontologists — determining exactly when extinction events began and ended can help to establish whether they were caused by a single catastrophic event or a slow change in the environment.

Geologists now want to repeat that process for other key events in the Earth's history. Last week's meeting, sponsored by the National Science Foundation (NSF), was held to start figuring out how to do this. The data would complement other NSF-funded geoinformatics projects already under way, including Chronos, which aims to match different layers of rock of the same age from around the world, and the Paleobiology Database (see *Nature* 424, 482–483; 2003), which looks at fossil finds.

Researchers are now putting together a proposal for how this project can be tackled.

Bowring believes that three new US labs will be required, at a cost of around \$6 million and with annual operating costs of between \$2 million and \$3 million.

"It's in everyone's best interest, so there's no reason why we shouldn't be able to get agreement," says Tony Fallick, who heads isotope dating at the Scottish Universities Research and Reactor Centre in Glasgow. Although the initiative is US-led, the lab work and results would be shared internationally.

The NSF has a new funding stream dedicated to efforts too big for one lab to handle. Geoinformatics is one such area, says Walter Snyder, section head for research grants in the NSF's Earth science division. "The burden is now on the community to agree on a common approach," he says.

Researcher faces life in prison for revenge radiation poisoning

Beijing A Chinese medical researcher has received a suspended death sentence for attempting to poison a business rival with radioactive material. Such sentences usually result in life imprisonment.

Gu Tianming and the victim, identified in government news accounts only by his surname, Liu, had run a laser-treatment centre together but had apparently fallen out over business dealings. In an act of spite, Gu forged documents to buy an industrial machine that uses iridium-192 — a radioactive element that anti-terrorism experts fear could be used in a 'dirty bomb' to attack populated areas. He then secreted pellets of the element in the ceiling tiles of his colleague's office in a hospital in Guangzhou. Liu suffered fatigue, headaches and vomiting; a further 74 people in the hospital, including a pregnant woman, also reported symptoms.

Gu's assistant, Fang Zhenhua, was sentenced to 15 years in jail for helping Gu with the plan.

Top lab contracts awarded as US battles deadly bugs

San Diego Massachusetts and Texas have won the top prizes in the US National Institutes of Health's competition for contracts to build high-security biodefence research labs (see *Nature* 425, 110; 2003).

Boston University Medical Center and the University of Texas Medical Branch at Galveston will each be given \$120 million to build the National Biocontainment Laboratories. The facilities will house labs rated as biosafety level 4, and so will be able to handle exotic agents that pose a high risk of life-threatening diseases, for which there is often no vaccine or therapy — such as Ebola, hantavirus infections and Rift Valley fever.

Nine other universities will receive between \$7 million and \$21 million to establish Regional Biocontainment Laboratories, which will examine less dangerous agents.



Handle with care: the United States is to build labs to house the world's deadliest pathogens.

Climate model gets whirlwind treatment

London A state-of-the-art climate model is now running on one of the world's fastest computers.

The Earth Simulator supercomputer (see *Nature* 416, 579–580; 2002) in Yokohama, Japan, can perform more than 30 trillion operations per second, and will be used to run programs investigating everything from geosciences to weather forecasting. The computer's current task is to run one of the world's most advanced climate models, developed by the Hadley Centre for Climate Prediction and Research in Bracknell, near London. The model can be run at a much higher resolution on the supercomputer: the atmosphere is divided into 100-kilometre squares, rather than the 300-km squares typically used by the Hadley Centre's own computer. This will allow researchers to capture small-scale weather features, such as hurricanes (pictured), that other models might miss.



"This makes a huge difference to what we can do," says Julia Slingo, a climate modeller at the University of Reading, UK. Results from the Earth Simulator, which went online in March 2002, were presented at a workshop on Earth-system modelling in Cambridge, UK, last week.

Ecstasy proposed as help for stress victims

Washington The 'clubbers' drug' ecstasy may be put to a clinical use, now that claims of deadly side-effects have been retracted. Experts at the Multidisciplinary Association for Psychedelic Studies in Sarasota, Florida, think that the drug could be used to treat post-traumatic stress disorder.

The study was approved by the US Food and Drug Administration in November 2001, but the team has struggled to find an independent review board that is prepared to monitor the study. The project received a near-fatal blow last year when another group claimed that low doses of the drug killed several laboratory monkeys. But their findings were retracted last month after the researchers realized that they had dosed the monkeys with methamphetamine — also known as 'speed' — instead of ecstasy.

In the wake of the retraction, an undisclosed board agreed on 23 September to monitor the stress study. The researchers now have one more hurdle to clear: they need a permit from the US Drug Enforcement Agency to dispense the pills.

Search comes to an end as Cell fills editor's chair

Washington A new editor takes the helm of the prestigious molecular biology journal *Cell* this week. After a prolonged search, Cell Press has appointed Emilie Marcus, previously editor of *Neuron*, another of the publisher's titles.

Cell's previous editor, Vivian Siegel, left in January to head the Public Library of Science, an open-access publishing venture based in San Francisco (see page 554).

Cell was founded by Benjamin Lewin in

1974. He propelled it to international renown and remained as editor until 1999.

Marcus has pledged to maintain close contact with researchers, and always be available to discuss manuscripts with authors, as Lewin was. "We will have a very accessible, open office," she says.

Help basic science, cry European Nobellists

Brussels As the Nobel Prizes were being announced in Stockholm this week, a group of 45 European Nobel laureates was busy lobbying in Brussels. They handed the European research commissioner, Philippe Busquin, a letter urging him to assign highest priority to establishing a European Research Council, a proposed agency for funding curiosity-driven research across all disciplines (see *Nature* 419, 108–109; 2002).

The letter's signatories argue that current European research programmes are not suited to basic science. A flexible, 'bottom-up' agency, driven by researchers rather than administrators, is an essential component that is missing from Busquin's vision of pan-European research, they say.

"There is an overwhelming level of support for the European Research Council among Europe's Nobel community," says Erwin Neher, a neuroscientist at the Max Planck Institute of Biophysical Chemistry in Göttingen, Germany, and joint winner of the 1991 Nobel Prize in Physiology or Medicine, who initiated the letter.

Correction

The News Feature "Spare me the lecture" (K. Powell *Nature* 425, 234–236; 2003) should have credited a team led by Daiyo Sawada and Michael Piburn of Arizona State University with developing the Reformed Teaching Observation Protocol.



Campaign trail: the Public Library of Science's founders (above) have unleashed posters and a television advert (see stills below) to publicize their project.

Who will pay for open access?

A new biology journal, positioned to compete with the likes of *Nature*, *Science* and *Cell*, aims to reinvent the economics of high-quality scientific publishing. Declan Butler examines the bottom line.

In June and July, television viewers in three US cities were treated to a 30-second commercial, in which a besuited man emerged from his house, to the voiceover: "In the year 2003, the Public Library of Science made it possible for people all over the world to have access to the latest scientific discoveries. Shortly thereafter, things began to change." In a scene reminiscent of *The Matrix: Reloaded*, the man then zoomed off into the sky.

To the average viewer, it must have been perplexing stuff. The event being promoted was this month's launch of a journal called *PLoS Biology*, which aims to publish articles of "exceptional significance" in disciplines from molecular genetics to ecology. Published by a non-profit body called the Public Library of Science (PLoS), it is the spearhead of a campaign to provide open access to the scientific literature—making it free for anyone to read.

It's a compelling idea. Imagine, for instance, that you have just been diagnosed

with a life-threatening illness, and want to find out about the latest advances in treatment. A search of the Medline literature database throws up hundreds of pertinent research papers. But unless you have subscriptions to the journals in question, or rack up your credit-card bill to download individual articles, many of the full texts will remain out of bounds—even though your taxes helped to pay for much of the research.

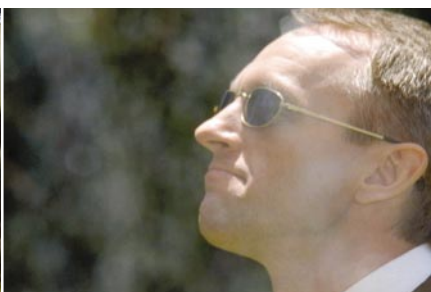
PLoS's aim is to show that open access can work by competing head-on for the best research papers with today's top scientific and medical journals, such as *Nature*, *Science*, *Cell* and *The New England Journal of Medicine*. *PLoS Biology* will be joined next year by *PLoS Medicine*, and perhaps later by a series of discipline-specific publications.

Few people would disagree, in principle, with the ideal of open access. The question is whether the economics can be made to work. Employing peer review to sift through hun-

dreds of manuscripts, and then editing the accepted ones into shape, can be an expensive business. Conventionally, publishers recoup these costs—and in the case of commercial publishers, make some of their profits—by charging for access to the final product. PLoS aims to turn this 'reader-pays' model on its head, instead charging a 'dissemination fee' to the authors of accepted papers. It believes that US\$1,500 per published article will do the job. In practice, the expectation is that this will be paid by universities or institutes, or by the agencies that fund researchers' work.

Behind the rhetoric of PLoS's campaign, therefore, lie some more down-to-earth questions. Will scientists, their host institutions, and those who fund their research embrace the 'author-pays' model? And if they do, is \$1,500 per article enough to cover the costs of producing a journal of the highest quality?

For most researchers in the physical sciences, PLoS's campaign is a side issue: they



routinely make their papers freely available, before formal publication, using online preprint archives such as arXiv.org. But for biologists, who are not generally comfortable with prepublication, the answers to the questions thrown up by the launch of *PLoS Biology* may define the future of scientific communication. How things will pan out remains unclear. "We'll be watching with interest to see whether PLoS and others can make the economics work," says Annette Thomas, managing director of Nature Publishing Group.

PLoS began life in 2000, when a group of leading biologists circulated an open letter requesting that scientific publishers make the contents of their journals free for all to access online within six months of initial publication. PLoS's founders included Nobel laureate Harold Varmus, president of the Memorial Sloan-Kettering Cancer Center in New York, DNA-microarray pioneer Patrick Brown of Stanford University in California, and computational and evolutionary biologist Michael Eisen of the Lawrence Berkeley National Laboratory, also in California. Their letter attracted more than 30,000 signatures — although few signatories seem to have followed through on their pledge to stop submitting to and reviewing for journals that have not acceded to PLoS's call for open access. These journals remain in the majority — hence PLoS's decision to launch its own publishing enterprise.

Mixed strategy

PLoS is not alone in exploring the possibilities of open-access publishing. Several scientific publishers, such as Oxford University Press and Britain's Institute of Physics, are experimenting with open access. Some journals, such as the American Physiological Society's *Physiological Genomics*, are allowing authors to pay for open online access for individual papers, while retaining a subscription model for the journal as a whole. And 1999 saw the launch of the London-based BioMed Central (BMC), which today boasts a roster of almost 150 online open-access journals, most of which have now published their first handful of papers.

In its pure, open-access form, *PLoS Biology* will also be an online product — a print version will be available for \$160 per year. But PLoS differs from BMC in its lofty editorial ambitions. It has pitched itself at the elite end of the market, competing with journals that



Universities may be reluctant to fund open-access papers while still paying for library subscriptions.

aim to publish only the best or most significant research. This can mean rejecting more than 90% of submitted manuscripts, which in turn means that editorial costs per published paper are very high — much higher than \$1,500, say some experts.

Many publishers decline to reveal details of their costs. But most have examined the issue and find PLoS's estimate hard to understand. Publishing primarily online should reduce production costs by around 20%, compared with a journal with a large print run. But that still leaves considerable costs for staffing and administration. "I feel that PLoS's estimate is low by four- to sixfold," says cell biologist Ira Mellman of Yale University, editor of the *The Journal of Cell Biology*, published by Rockefeller University Press. He notes that PLoS has hired six full-time editors — some former employees of *Nature* or *Cell* — and is based in San Francisco, an extremely expensive city.

PLoS's supporters reject such arguments, pointing out that they come from journals with a vested interest in the reader-pays status quo. Eisen says that the \$1,500 figure is based on a 100-page business plan thrashed out with publishing experts. "PLoS needs to break even in just a few years, so we have every incentive to get the number right," he says.

In the first instance, PLoS will be cushioned by a grant from the San Francisco-based Gordon and Betty Moore Foundation. If PLoS meets publishing targets agreed with the foundation — which have not been made public — this grant will total \$9 million over five years.

But in the longer term, PLoS and other open-access groups must persuade the organizations and institutions that fund and host biology research to pay their fees. Grants from

the US National Institutes of Health already allow for the charging of publication fees. Other bodies are moving in the same direction. The Howard Hughes Medical Institute, for example, has agreed to provide its investigators with up to \$3,000 each in 2004 to cover open-access fees. On 1 October, Britain's Wellcome Trust announced that it, too, is prepared to meet the costs of open-access publishing (see *Nature* 425, 440; 2003). And Germany's main research agencies are expected to make a similar announcement later this month.

Transition time

PLoS and other open-access groups are now waiting for word from the bulk of universities and research institutes around the world. "We are in a transitional period," says Peter Suber of the advocacy group Public Knowledge, based in Washington DC. He notes that many institutions will be reluctant to cover dissemination fees while still paying subscriptions to traditional journals.

Before these issues are resolved, however, *PLoS Biology* will have to overcome the hurdle that faces any new journal, irrespective of its business model: convincing scientists, particularly young researchers who need to publish in high-profile journals to further their careers, that it is worth taking the risk on a new and unknown quantity. "We are off to a very fast start, with submissions steadily increasing even before we've launched," claims Eisen.

Most publishers remain sceptical about the viability of PLoS's eventual goal of converting the entire scientific literature to the open-access model. But many now accept that the author-pays approach may have its place. In August, the Association of Learned and Professional Society Publishers declared itself "wholly in favour of maximizing access to research literature; the various proposals for achieving this ... raise complex economic, logistical and sociological questions which differ from field to field as well as between different sizes and types of publishers. Much more information needs to be gathered through experimentation and analysis."

With the launch of *PLoS Biology*, a key experiment is now up and running. ■

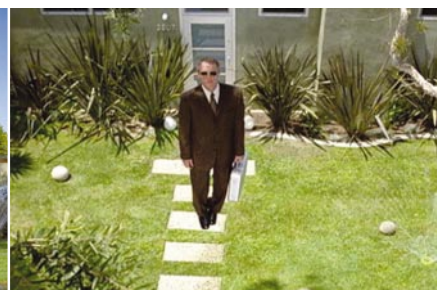
Declan Butler is *Nature's* European correspondent.

Public Library of Science

♦ www.plos.org

BioMed Central

♦ www.biomedcentral.com



Think outside the sandbox

One grain of sand is a solid. But a lot of grains together can behave like a solid or a liquid. By probing this dual personality, physicists hope to understand a host of real-world systems, says Mark Buchanan.

Heinrich Jaeger knows what it's like, literally and metaphorically, to watch the sand slipping through the hourglass. "We thought it would be a four-week experiment," the physicist recalls of one of his studies. "It ended up taking several years."

A decade ago, Jaeger and his colleagues at the University of Chicago were busy studying what happened when they poured sand into a container and tapped the box repeatedly. They expected the grains to bed down fairly quickly — but, as the team eventually reported in 1995, the sand kept shifting into ever more compact configurations¹. "Granular systems look deceptively simple," says Jaeger. "But the closer we look, we find so many complexities."

Such puzzles show why granular materials hold a special fascination for physicists interested in the fundamental properties of matter. A pile of sand, flour or mustard seed can stand tall and fixed like a solid, defying the tug of gravity that would bring water sloshing to the floor. Yet sand pours through an hourglass as though it were a liquid.

People spend more time and effort in handling granular matter than any type of material aside from water. In industry, powders clog chutes and otherwise misbehave, at enormous economic expense. So understanding how simple forces — mostly gravity and friction — help to shape the materials' multiple personalities is of immense practical significance.

Out of equilibrium

What's more, some physicists believe that progress towards a theory of granular matter could provide clues to the wider world of 'non-equilibrium' physics. This is a large missing piece of the theoretical framework that physicists use to describe matter, and applies to the vast majority of real-world systems, which are not held in fixed and unchanging conditions but instead are shaken, stirred, heated, or driven by energy in countless other ways.

Theorists have long approached granular matter by considering it to be analogous to ordinary matter, with the grains playing the roles of individual molecules². At rest, a pile of grains is indeed rather like a solid, able to resist a certain amount of stress. But if you shake it vigorously, letting the grains tumble past one another and rearrange themselves, it does appear to shift into a



Stick or slick? Sand can be sculpted into solid forms, both by humans (above) and by the elements (above right). But put it in an hourglass and it flows like a liquid — and theorists want to know why.

'liquid' form. Much theorizing about the behaviour of granular matter therefore centres on the extent to which it is meaningful to talk about 'melting' piles of sand, and whether the degree of shaking might be related to an 'effective temperature' of the granular liquid.

Recent experimental results have revealed some fundamental similarities between granular liquids and their conventional counterparts. High-school physics students are familiar with the concept of brownian motion, in which a small particle floating in a pool of liquid dances around, driven by the thermal motions of the surrounding molecules. And Albert Einstein showed that the speed with which the particle wanders is linked to how easily the same particle can be dragged through the liquid by an external force³. Two numbers — a 'diffusion coefficient', D , and a 'mobility', μ — describe these two behaviours, and their ratio is always equal to the temperature of the liquid, T .

In August this year, Gianfranco D'Anna and his colleagues at the Swiss Federal Poly-

technic in Lausanne showed that this relationship also holds true for granular liquids⁴. They put 50,000 glass beads in a small container, which they vibrated at high frequency to generate 'liquid' behaviour. The team used this set-up to run a simplified version of Einstein's experiment: whereas he considered a particle that could wander freely, the cone-shaped 'particle' used by D'Anna's team was able only to rotate.

Go with the flow

Whatever the size of the cone, the ratio between D and μ was constant. And this effective temperature seemed to reflect accurately the vigour of the disordered motion of the constituent particles — just as a thermal temperature does for an ordinary liquid. "A granular liquid is indeed a fairly good liquid," says D'Anna.

The team's results lend support to the idea that granular matter is not such a misfit after all. They also raise an obvious question: if adequate vibration generates a liquid, what happens if the vibration is less strong? Weaker vibration corresponds to a lower

move, but with great difficulty. Because these structural rearrangements occur so slowly, the material is stable for most practical purposes. However, the physical properties of the glass — such as its electrical conductivity — gradually change as its structure evolves. Ordinary matter in equilibrium does not experience this sort of ‘ageing’.

Since Jaeger’s team reported its first results, theories developed to describe both granular solids and glassy materials have begun to converge — suggesting that the two types of matter are fundamentally similar.

Grains of truth

In the late 1980s, Sam Edwards of the University of Cambridge, UK, tried to formulate a theory of granular matter for one special case — a dense collection of grains flowing just fast enough to avoid sticking. For this regime, Edwards proposed that an accurate theory could be devised by supposing that the grains are equally likely to fall into any of the possible ‘jammed’ configurations³.

At the time, most physicists were unconvinced by the theory — it seemed possible, if not likely, that some of the jammed configurations would be more probable than others. “It was a very bold proposal,” says Jorge Kurchan of the Ecole Supérieure de Physique et de Chimie Industrielles in Paris. “Few took it very seriously.” A theory as simple as Edwards’, physicists reasoned, was unlikely to explain the complexities of granular behaviour.

By the late 1990s, however, Kurchan and other theorists were developing ways to understand ageing in glasses in fundamental terms⁶ — and their theories bore a striking resemblance to Edwards’ theory of granular matter. “In retrospect, his ideas were not so crazy after all,” says Kurchan.

Last year, Kurchan teamed up with Hernán Makse of the City College of New York to test Edwards’ ideas directly. Using a computer, they simulated the flow of grains trapped between a fixed lower surface and an upper surface moving just fast enough to keep the grains flowing. Like D’Anna and his colleagues, Kurchan and Makse measured the effective temperature of the grains by examining the relationship between D and μ for a simulated probe particle. Kurchan and Makse also calculated the effective temperature using Edwards’ theory, and found that the two results matched⁷.

Nonetheless, a full theory of granular matter remains some way off. As Kurchan points out, Edwards’ ideas do not apply to

granular liquids, in which the grains leave their jammed configurations behind. And some physicists wonder about the utility of the ‘effective temperature’ concept. “It is not clear,” says Douglas Durian of the University of California, Los Angeles, “if these temperatures have a predictive significance.” In earlier experiments, for example, effective temperatures measured for different types of flow tended to have different values, whereas a truly fundamental temperature should give the same value no matter how it is measured.

But the fact that some of the same fundamental ideas seem to apply to both granular matter and glassy materials is encouraging, as both are examples of non-equilibrium systems — and gaining a general understanding of these systems is near the top of the ‘to do’ list for theoretical physicists.

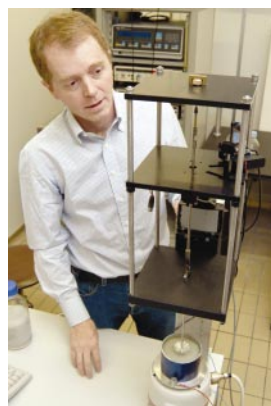
Current theories of solids, liquids and gases are based on equilibrium statistical mechanics, which apply only to systems that are in thermal equilibrium. A bowl of water at room temperature is one such system, as it has the same temperature as its surroundings. But a granular liquid is not in equilibrium — it needs a constant input of vibrational energy to retain its ‘liquid’ state. A ‘jammed’ granular system is also out of equilibrium, as the grains cannot easily explore all configurations. Similarly, the way in which the character of glass evolves with time is a direct consequence of the system having excess energy and taking an

incredibly long time to reach equilibrium.

The problem, for physicists interested in the fundamental behaviour of matter, is that most real-world systems are not in equilibrium. From the chemistry of living cells to the river of cars on a crowded highway, the real world is out of balance, perpetually flowing and evolving. If the recent surge of interest in granular matter leads to a deeper understanding of non-equilibrium systems in general, the theoretical pickings could be rich indeed. “Intellectually,” says Jaeger, “things are beginning to gel.” ■

Mark Buchanan is a freelance writer based in Cambridge, UK.

1. Knight, J. B., Fandrich, C. G., Lau, C. N., Jaeger, H. M. & Nagel, S. R. *Phys. Rev. E* **51**, 3957–3963 (1995).
2. Jaeger, H. M., Nagel, S. R. & Behringer, R. P. *Rev. Mod. Phys.* **68**, 1259–1273 (1996).
3. Einstein, A. *Ann. Physik* **17**, 549–560 (1905).
4. D’Anna, G., Mayor, P., Barrat, A., Loreto, V. & Nori, F. *Nature* **424**, 909–912 (2003).
5. Mehta, A. & Edwards, S. F. *Physica A* **157**, 1091–1097 (1989).
6. Cugliandolo, L. F., Kurchan, J. & Peliti, L. *Phys. Rev. E* **55**, 3898–3914 (1997).
7. Makse, H. A. & Kurchan, J. *Nature* **415**, 614–616 (2002).



Gianfranco D’Anna has studied the physics of ‘granular liquids’.

effective temperature. So, by analogy, a progressively less vigorously shaken granular liquid should reach a point at which it ‘freezes’ into a solid.

These are the conditions explored by Jaeger and his colleagues in their studies during the early 1990s. But their grains did not freeze into anything like an ordinary solid. Contrary to expectations, a few days of tapping did not bring the container of sand into a fully compact state. In some cases, the grains remained unsettled even after several hundred thousand taps¹. As Jaeger’s group suggested, this is similar to the way in which glassy materials solidify.

Upon cooling, the molecules of a glassy material do not fall into a regular, crystalline arrangement, but instead become jammed in a disordered structure. The molecules still

'Open access' will not be open to everyone

Making authors pay for publication may not deliver the anticipated benefits.

Sir — With the best intentions, a group of American biomedical scientists and physicians proposes to make all scientific research “available free of charge to anyone”. They have formed the Public Library of Science (PLOS) and plan to build their library by starting new journals, all of which will be open access. They will support their journals by charging authors, not subscribers, for each paper that is published. The fee will be something like US\$1,500 per paper — maybe more, maybe less. Their model, they say, “treats the costs of publication as the final integral step of the funding of a research project”.

In reality, not all researchers are funded by research grants. Less than half of active research mathematicians are supported by federal grants in the United States; some estimates are considerably lower than this. Mathematics is not alone: in virtually every field, there are many scientists doing outstanding research who are not part of any large, federally funded project.

Who pays the \$1,500 for these people?

Much published research also comes from outside the United States. More than half of the articles published in the American Mathematical Society's primary research journals during the past 20 years do not have a US author. Who pays the \$1,500 for these authors? Yes, it's true that many such authors come from countries that have funding agencies themselves, but many others are from developing countries. During those same 20 years, of the 1.5 million journal articles listed in *Mathematical Reviews*, 68,000 had authors from the former Soviet Union, 26,000 from India, 26,000 from Poland, nearly 4,000 from Egypt, and more than 600 from Nigeria. Who pays the \$1,500 for these scientists?

The PLOS editors suggest that authors who are unable to pay won't have to. But this assumes that few authors are unable to pay — a false premise in many disciplines.

It is often assumed that universities or

institutes will pay the \$1,500, if no one else will. But how will universities and departments decide which faculty and which areas of research are supported? What happens to faculty in small colleges with limited resources? Publishing papers only from those who can pay is a dramatic change in culture, and the change may not be the intended one.

Each publication model — subscription-based or author-supported — has trade-offs, but they are not symmetric trade-offs. When a scientist doesn't have a subscription, he or she can nonetheless get information about the article (the abstract and perhaps a list of references); requesting a copy of the article can be as easy as sending an e-mail. When a scientist doesn't have the funds to publish an article, the article does not appear — does not become part of the permanent literature. That's more than an inconvenience.

John Ewing

American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA

Flaws undermine results of UK biotech debate

Sir — The *GM Nation?* report released in the United Kingdom last month (see *Nature* **425**, 331; 2003) concluded that the general public is overwhelmingly against genetic-modification (GM) technology, with feelings ranging from “suspicion and scepticism, to hostility and rejection”.

The study cost £500,000 (US\$830,000). Unfortunately, this was not money well spent. The methodology was so badly flawed that the data not only failed to support the authors' conclusions, but undermined them.

The main fault with the study, as the authors concede, is the self-selected nature of the main sample. About 36,000 people took part in an “open debate”: half of these responses came by mail and half from the *GM Nation?* website (www.gmnation.org.uk). This sample is certainly large, but it is not random. It is in fact most likely to attract those who have strong opinions about GM. One might think that the sheer size of the sample swamps any problems with its self-selected nature, but for that to be the case you would need millions of participants.

Although the authors of the report were aware of this criticism, they offered only two countermeasures. First, they checked a random sample of responses to see if there were any standardized ones being sent in by activist groups, which there weren't. But

people with strong views on GM are capable of expressing their own opinions, and this measure does nothing to prevent the sample being biased in favour of them.

Second, they commissioned a ‘narrow-but-deep’ study from another company, “as a control on the self-selecting participants in the open debate”, to see if there was a “silent majority” with different views. This meant asking 78 people 13 questions from the open debate. This sample was randomly chosen — although the report is short on specific details — and stratified so that it roughly matched the general population. (This group was also re-tested after 2 weeks of group discussion and personal research to see if their attitudes to GM changed.)

The authors of the narrow-but-deep section conceded that their results were not statistically robust, because of the small numbers involved. Nevertheless, they said: “We believe it is an accurate reflection of the general public.” The initial responses of the random group were, however, noticeably different from the results of the open debate. (Even after 2 weeks, the differences, although not as large, remained significant.) Yet the *GM Nation?* report claimed that, apart from some minor differences, the two groups agreed. The general public, said the authors, is not “a completely different audience with different values and attitudes from an unrepresentative activist minority”.

The actual results from the two groups were buried within the supporting

documents, far apart from each other. Once these results are compared side-by-side, startling differences emerge for more than half of the questions used (visit www.nottingham.ac.uk/philosophy/staff/Campbell/Table1.htm for a full comparison).

For example, to the question “I would be happy to eat GM food”, only 8% of the open-debate respondents agreed, compared with 35% for the random group. On the topic of whether GM was unnatural, 84% thought so in the open debate, but only 37% did in the random group.

We find it astonishing that the obvious mismatch between the random group and the open-debate group was not discussed anywhere in the report, and that it did not prevent this report being released and becoming headline news.

With £500,000, a larger version of the narrow-but-deep study could have been conducted, avoiding the problem of self-selection. As well as using a ‘topic blind’ recruitment strategy, questions about GM food would ideally be embedded among questions about other current concerns, so that the participants would be unaware that GM food was the focus of the research. Also, the sort of vague and leading questions used by *GM Nation?* should be avoided. Only then could we be confident that the findings are reliable and realistic.

Scott Campbell, Ellen Townsend

Institute for the Study of Genetics, Biorisks and Society (IGBiS), University of Nottingham, University Park, Nottingham NG7 2RD, UK

Mad, bad and dangerous to eat

What the public wants remains a mystery, but food safety can't be ignored.

When Food Kills: BSE, *E. coli* and Disaster Science

by Hugh Pennington

Oxford University Press: 2003. 226 pp.

£25, \$35

Eileen Rubery

There is a need for a book that explains to the public what food safety is all about and that gives policy-makers and managers the conceptual framework to make sensible judgements on 'food crises' when they hit the headlines. The preface of this book does not say for whom it is written or what its aims are. But it comes from Hugh Pennington, who led Britain's inquiry into the outbreak of food poisoning caused by the bacterium *Escherichia coli* O157 in Wishaw in Scotland in November 1996.

Sadly, although it is good in parts, the book does not explain the basis of food-safety assessment to the generalist. It lacks coherence and fails to develop clear lines of argument and explanation. Neither is it helpful to public-health specialists, general microbiologists or crisis managers, as it gives no advice that they could follow, nor any proposed framework for decision-making. It left me no clearer as to what Hugh Pennington thinks government or local authorities should do when faced with food scares, nor what he feels the public should think.

Instead, the book is a *mêlée* of good stories and fascinating anecdotes from the world of public health, loosely strung together around the stories of 'mad cow disease' (bovine spongiform encephalopathy, or BSE), new variant Creutzfeldt–Jakob disease (vCJD) and *E. coli* O157. The Wishaw food-poisoning outbreak was caused by lax hygiene in a local butcher's shop: 'ready-to-eat' products were contaminated with *E. coli* O157 from raw meat, and many people died. Occurring nine months after the BSE crisis, when the Conservative government had a fragile majority, the Pennington report made a number of recommendations, most of which have now been implemented.

But in *When Food Kills*, Pennington does not mention what to me was the really surprising outcome — that the local population sympathized with John Barr, the butcher involved, even when it became clear that he continued to distribute meat products after being identified as the source of the outbreak, thus causing many additional cases.

Why was the public so apparently illogical? Similar situations have arisen elsewhere, notably in connection with the risk of meningitis and permanent brain damage in the unborn fetus if a pregnant woman eats



Condemned: John Barr sold lethally infected meat, but local people still supported him

cheese contaminated with *Listeria* and contracts listeriosis. Another example is the public's negative response to the banning of 'beef on the bone' when concerns arose about BSE prions in lymph nodes. In all these cases, it seems that the public saw heavy-handed enforcement, not an attempt to protect them. As a former policy-maker in the area of food safety, I know how difficult it is to decide what the public will perceive as the correct balance between the freedom to ply a trade and the need to protect the public.

Since the Pennington report was published, Britain's Labour government has tried to address public concerns about food safety by setting up the 'independent' Food Standards Agency (FSA). This has had some initial success (see *Nature* **410**, 867–868; 2001), although without being tested by a really big outbreak, it is probably too early to judge its performance. Food safety is a perennial obsession of the media, which tend to focus on food crises roughly every ten years: *Salmonella* in eggs and botulism from hazelnut yoghurt were key concerns around 1986, then interest fell away until BSE and *E. coli* O157 in 1996. I predict that 2006 or 2007 will be the critical year for the FSA, and hope they are doing some forward planning.

Another problem with *When Food Kills* is that Pennington starts his stories from the end (the cases of illness) and works back to the beginning. So we hear about vCJD victim Stephen Churchill developing an unsteady gait, about the Foré people dying from the prion disease kuru because they eat their ancestors' brains, and about the use of ultracentrifuges to show that the infectious agent in prions is unlikely to be a nucleic acid. Only then are we told about BSE in cattle and the facts related to the BSE epidemic. This leads

to a confused feeling of 'so what', rather than 'so now I understand'. A number of regulatory and inspectorial issues provide some context but few conclusions. We range as far afield as the Three Mile Island nuclear accident, train crashes from the distant and recent past, and problems in old lunatic asylums, but no strong theme or message comes from it all.

I did enjoy his quotes from what must otherwise be dry reports of public inquiries. For example, when talking of the *Salmonella* food-poisoning outbreak at Wakefield's Stanley Royd Hospital in 1984, Pennington reports that the kitchens had been recognized as hazardous for many years, but that the public inquiry concluded that planning activities were a "remarkable example of what well-intentioned individuals can fail to achieve unless someone is charged with the responsibility of ensuring that careful attention to detail does not lead to a complete cessation of all activity other than the production of paper". Unfortunately, today's well-intentioned emphasis on consultation and team-working probably exacerbates, rather than ameliorates, this problem.

Then Pennington quotes from the report of the inquiry into the 1966 Aberfan landslide disaster: "We found that many witnesses ... had been oblivious" to the dangerous state of the tips. "It did not enter their consciousness. They were like moles being asked about the habits of birds."

Individually these stories are entertaining and provide memorable examples of the human propensity to ignore danger signals until it is too late. Indeed, one ends up surprised that disaster did not strike far earlier, or more often, given the sagas of negligence and misunderstandings detected once an

investigation begins. But the overall impact is reduced by the lack of critical analysis of the cases, and unfocused and superficial discussion of the difficulties of keeping alert those charged with protecting us from rare but serious dangers.

Although overall the book disappoints, I know I will be using Pennington's quotes in my lectures in future. Indeed, one quote from the preface will remain with me. In Shakespeare's *Twelfth Night*, the foolish knight Sir Andrew Aguecheek anticipates the BSE crisis by saying "but I am a great eater of beef and I believe that does harm to my wits," to which Sir Toby, his companion, replies: "No question." The bard, as ever, was far ahead of his time. ■

Eileen Rubery is at the Judge Institute of Management, University of Cambridge, Trumpington Street, Cambridge CB2 1AG, UK.

Prime time for mathematics

Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics by John Derbyshire

Joseph Henry: 2003. 456 pp. \$24.95, £19.95

The Music of the Primes: Why an Unsolved Problem in Mathematics Matters

by Marcus du Sautoy
Fourth Estate: 2003. 335 pp. £18.99
HarperCollins: 2003. \$24.95

W. T. Gowers

For many years, the Riemann hypothesis has been regarded as one of the most important unsolved problems in mathematics. If it could be proved, it would go a long way towards demonstrating what observation already suggests — that in many ways prime numbers are distributed as if they had been chosen randomly. Popular science is big business these days, and there is a rapidly developing sub-genre of books about famous mathematical problems, either unsolved or recently solved. So it was inevitable that attention would turn to the Riemann hypothesis, especially as it now carries a bounty of a million dollars, offered by the Clay Mathematics Institute. And, sure enough, now there are three popular books on the subject: *Dr. Riemann's Zeros* by Karl Sabbagh (Atlantic: 2002) and the two reviewed here.

Unfortunately, and in contrast to other mathematical problems such as Fermat's last theorem, Kepler's conjecture and the travelling-salesman problem, the basic statement of the Riemann hypothesis — that the non-trivial zeros of Riemann's zeta function all have real parts equal to one half — is rather sophisticated and difficult to explain to a

general audience. The zeta function is defined by the infinite series

$$\zeta(s) = 1^{-s} + 2^{-s} + 3^{-s} + 4^{-s} + \dots$$

so it is important to understand when a series such as this makes sense. Next, we need to understand Riemann's crucial innovation, which was to extend the definition of the zeta function to include complex numbers (the function had already been considered by Euler for real s): even readers who are familiar with i , the square root of -1 , may be baffled by the idea of 2^{-i} . After coming to terms with all that, one learns that the expression for the zeta function does not make sense when the real part of s lies between 0 and 1. Nevertheless, there is a way of making sense of the zeta function itself in this region, and, remarkably, its behaviour has important consequences for the distribution of prime numbers. Since it is the seeming randomness of the primes that lies behind much of modern cryptography, and in particular behind the security of the Internet, understanding these mysterious numbers is of the utmost importance.

In his book *Prime Obsession*, John Derbyshire sets himself the heroic task of explaining the Riemann hypothesis to readers who have no mathematical background beyond perhaps a basic fluency at rearranging bits of algebra and a half-forgotten exposure to calculus. Although all the steps are there, he is attempting the impossible: to learn a subject as hierarchical as mathematics, you have to understand and digest one level thoroughly before you can move on to the next. Indeed, he almost admits this in his introduction: "If you don't understand the Hypothesis after finishing my book, you can be pretty sure you will never understand it."

But it would be quite wrong to judge the book a failure for this reason. In common with almost all books of this kind, there are parts that you will skim, and which these are will depend on your mathematical background — but the bits you do read will be extremely well explained. A major and most unusual strength of the book, which even experts will enjoy, is a sort of intimacy between the author and the zeta function itself. Not content with the abstract definition, he tabulates values and draws numerous diagrams (not easy, as the four dimensions of the graph must be represented on a single page) to convey to the reader the intuitive feel for the

function that he himself quite clearly has.

Marcus du Sautoy's entertaining book *The Music of the Primes* is aimed at the more popular end of the market and looks certain to be a great success. He is less ambitious than Derbyshire in what he tries to explain, and is wider in his focus, covering prime numbers in general and not just the Riemann hypothesis. Although many of his anecdotes, such as what Ramanujan said about the number 1729, will be familiar to readers of other popular mathematics books, there are many other good ones that are less well known. I particularly recommend the enlightening story of how Don Zagier, a contemporary number theorist, changed his mind and started to believe that the Riemann hypothesis was true — and what that had to do with the most expensive bottle of wine ever.

The language used by du Sautoy is more hyperbolic than Derbyshire's — for example, the word 'stunning' makes regular appearances. He also likes to use certain metaphors repeatedly. Most notably, instead of saying, "the zeros of Riemann's zeta function all lie on the critical line", he prefers, "the points at sea level in Riemann's imaginary landscape all lie on the magic ley line". When such metaphors have been used often enough, they can end up simply as substitutes for more standard terms (and, indeed, many standard mathematical words were themselves chosen for their metaphorical suggestiveness).

Derbyshire has the same habit — for him, Euler's product formula is the "golden key", the turning of which is presented as one of the high points of Riemann's argument.

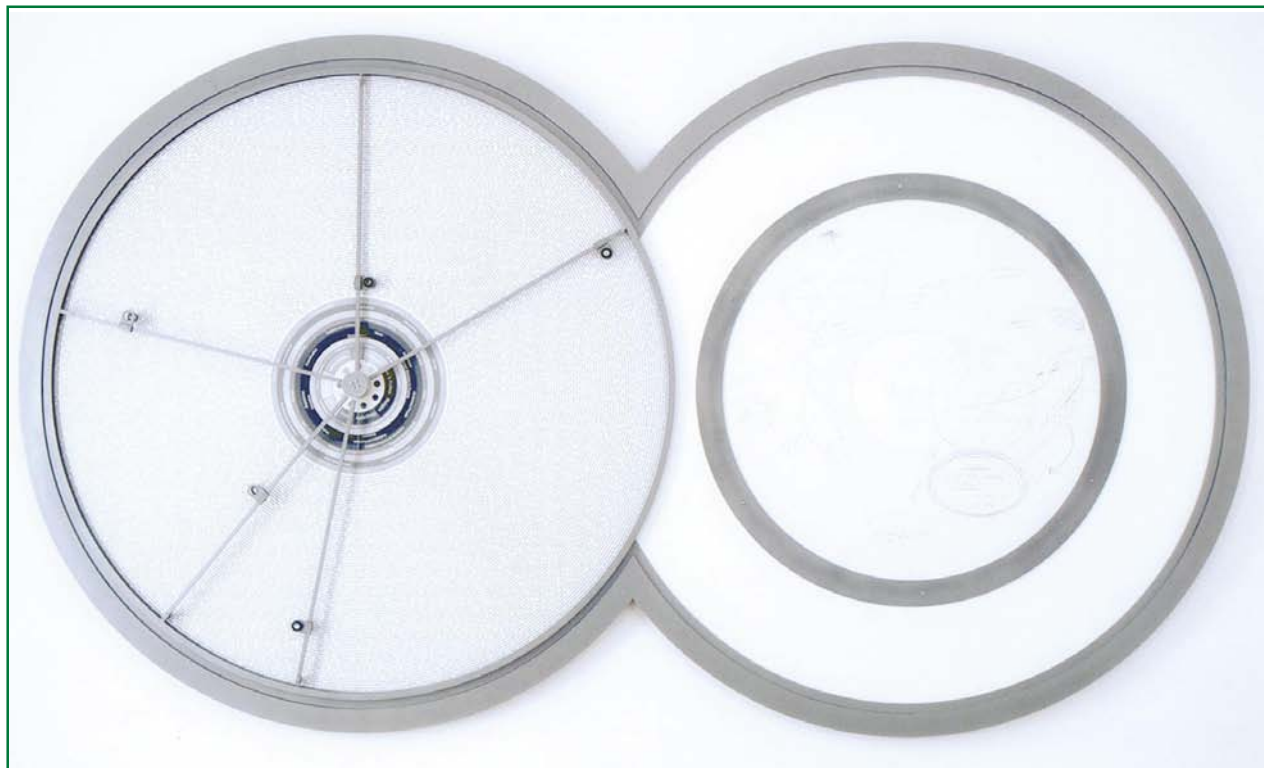
The chapters of Derbyshire's book alternate between mathematical ones and more historical ones, a device that works well. Riemann is one of the few undoubted geniuses of the subject, and had an extraordinary and all-pervasive influence on

mathematics. As well as the Riemann zeta function, he came up with the Riemann sphere, Riemann surfaces, Riemannian manifolds, the Cauchy–Riemann equations, the Riemann integral and numerous theorems that also bear his name. He was a shy man from a relatively humble background and much of his short life was a struggle against poverty, illness and depression. It is hard not to be moved that such a life left the world so enriched. ■

W. T. Gowers is at the Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK.



Science in culture



The art of rigorous chance

Keith Tyson has put a new spin on science in Berlin.

Martin Kemp

The diagram is the one genre of visual representation that may be credited to science — at least before the advent of modern techniques for seeing the invisible, such as X-rays. A diagrammatic image typically has a hard, impersonal, precise appearance, however personal the choices and errors that might have entered into its generation.

Unsurprisingly, few artists have made much of diagrams. The most notable exceptions were Paul Klee and Joseph Beuys — though Marcel Duchamp also did his bit. Klee, in the years between the two World Wars, became fascinated by the inherent dynamics of lines and colours in diagrammatic configurations, aspiring to devise a linear vocabulary that was more universal and less wholly symbolic than most diagrams.

Beuys, on the other hand, used diagrams as diagrams in his project of 'social sculpture'. In the many happenings and installations in which the shaman-like artist expounded his all-embracing ideas about science, technology, art, politics, economics and society, he smothered blackboards with chalk diagrams of his concepts. Many of his blackboards have been preserved as concrete records of otherwise intangible processes of thought and exposition — like Einstein's blackboard conserved in the Institute of Advanced Study at Princeton.

The impersonality and sense of process that

are potentially embodied in diagrams have recently attracted Keith Tyson, the British artist who is currently exhibiting his "Works from a Teleological Accelerator" at Arndt & Partner, in Berlin. Winner in 2002 of the most prestigious British art award, the Turner Prize, Tyson's 'manifestations' have consistently subverted expectations about art, the artist and the artwork. Looking into his studio, we are as likely to see boards covered in diagrams as normal artistic sketches. A residue of his initial training as a mechanical engineer is perhaps more apparent than the training he received in fine art.

Tyson first achieved international prominence with his *Artmachine*, which involves his programming a computer to access the vast bodies of information on the Internet, from which the computer generates proposals for the creation of artworks. The artist receives instructions, often incompatible and bizarre, for the making of a sculpture or painting, which he then obeys. The artist is the conceiver of the project, but not of the individual works, which arise from the impersonal results of an inexorable process. The *Artmachine* conducts its own creative experiments in chaos.

Tyson explains the relationship with the machine: "I program it, using language and making structures. And then it programs me [because] I make the work with my own hands, which forces me to reprogram the machine. ... It's a research project on myself. I'm seeing what are my limits ... I'm viewing it as if I'm experimenting on a frog."

In Berlin he has constructed the *Teleological Accelerator*, shown above: a large installation centred around a massive wall 'instrument', 5 metres wide. Two interlocking discs with moveable indicators permit the random combination and recombination of words and concepts covering many types of human endeavour and science. The 'direction' of the accelerator is characterized as predetermined, according to a teleological law, while the individual 'events' are random. It is as if evolution has been set on a path predetermined by what the physical laws of the Universe have decreed to be the optimum outcome, while working via a mechanism of natural selection involving chance combinations.

Tyson is not conducting science, obviously. What he is doing is to interweave the visual vocabulary and conceptual framework of science and technology with philosophical conundrums and the ironic questioning of artistic creativity. The results are not easy to appreciate and may lay themselves open to ridicule. But, in the face of the facile irony delivered by much contemporary art, the potential rewards of his work are considerable. *Martin Kemp is professor of the history of art at the University of Oxford and co-director of Wallace Kemp/Artakt.*

Keith Tyson's "Works from a Teleological Accelerator" is at Arndt & Partner, Zimmerstrasse 90-91, Berlin, Germany, until 25 October 2003.

Channel hopping

What was your first experiment as a child?

While riding with my grandmother in her 1939 Ford, I became fascinated with trying to figure out a way to estimate the total distance of our journey solely from tracking the angle of the speedometer throughout the trip. I never did come up with an algorithm for this, but perhaps that's why I was so excited by my first encounter with calculus in high school.

Who has been the most important mentor in your career?

Gilbert Ling, a brilliant but ultimately tragic figure whose theory that membranes don't exist appealed to the contrarian in me, then a clueless graduate student. Also influential was his inability to discard a theory becoming increasingly untenable in the inexorable flow of fact. This made me highly sensitive to Faraday's dictum about holding theories lightly on the fingertips, and of Scatchard's extension of that dictum to facts themselves.

What book has been most influential in your scientific career?

Gilbert Ling's *A Physical Theory of the Living State* (1962). In my final year as an undergraduate physics student, this book fired up my fascination with a biological question: why are Na^+ and K^+ handled so differently by cells when they are chemically so similar? If I had not stumbled upon this book, I might still be a high-school maths teacher, a job that I enjoyed very much before I returned to school for graduate work.

What literary character would you employ as a postdoc?

At first impulse, I thought to offer the job to Henry Burlingame III, the omniscient and personally appealing teacher in John Barth's comic masterpiece *The Sot-Weed Factor*. But on pondering this, I realized that Henry's experiments would all work perfectly the first time — his gels would be consistently without blemish and the single-channel records noise-free — and that consequently I'd always suspect his results to be fraudulent. So instead I'd choose razor-sharp, clear-eyed Dorothea Brooke from George Eliot's *Middlemarch*, who makes mistakes but always learns from them and is driven deeply by an uncompromised curiosity.

Does this mean that there's an ion-channel equivalent of the 'Key to All Mythologies'?

Yes. The role of the top quark in ion selectivity and gating — and we are all, at bottom, Casaubons.

What's your favourite conference destination, and why?

That's easy — Baltimore. It's simply a perfect place to exchange scientific information: a well laid-out conference hall close to cheap hotels, a surfeit of crab cakes, and not much in the way of distraction.

What was the most memorable comment you ever received from a referee?

"This paper will be very interesting to three people in the world and totally boring to everyone else." The reviewer (I think I know who you are — and where you live) had it exactly right. I was one of the three people.

What music heads the playlist in your car or lab?

In my car, Mose Allison's sublime jazz/blues piano. In the lab, the rule is no music composed after the death of Mozart. That generally takes care of things.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Sam Clemens. If a third were permitted to join us for brandy and cigars, I'd invite Henry Mencken.

What one thing would you rescue from your burning laboratory?

My daughter's first oil painting.

What's the best piece of advice you've received?

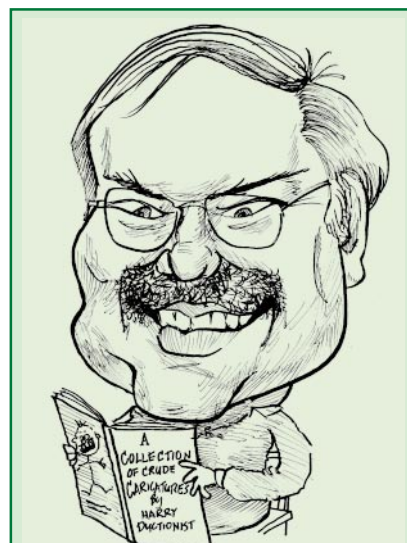
From my postdoctoral adviser Ef Racker, his exact words and urgent tone still vivid in my memory: "Don't think about the damned experiment — just do it!"

What do you most dislike about having research published?

The inevitable awkward waltz with the copy-editors, who, while slavishly following rules of syntax and diction learned as pre-adolescents (rules promulgated by their tin-eared employers), strive to disembowel my full-blooming paragraphs of every vestige of colour; my sinuous sentences of music; and my subtle words of connotative accompaniment. This problem is, paradoxically, far more prevalent in the popular vanity journals such as *Nature* than in workaday technical publications. But at least I'm not bitter.

Is there a 'tyranny of reductionism' in how scientists are trained today?

I don't accept the question's premise — that such a tyranny exists. The word 'reductionism' carries pejorative overtones that apply only to its crudest caricature, usually drawn by non-scientists, not to its actual implementation in the trenches. The caricature portrays us collecting free-floating facts about little meaningless things while missing the big picture on the



Chris Miller

Chris Miller is a Howard Hughes Medical Institute investigator and professor of biochemistry at Brandeis University in Waltham, Massachusetts, and works on ion-channel mechanisms.

pointillist canvas. But it seems to me that all great understandings of the big things emerge from meticulous attention to the detailed workings of the minutiae: solar systems from the investigation of real numbers, evolution from pigeon-breeding, and so on. Even the novel behaviours of complex systems arising from complexity itself are ultimately comprehensible only in terms of the underpinnings — as with Boltzmann's treatment of a huge collection of newtonian billiard balls vomiting up that humdinger of an emergent property, the second law of thermodynamics.

Do you have a burning ambition to do or learn something of no practical or immediate value? If so, what?

As I approach the age of 60, crossing the American continent on a bicycle has become the most appealing useless accomplishment I can muster in my imagination that is still marginally plausible.

The host at a dinner party has placed you next to God. What would you talk about?

I would be very curious to ask God: "On those rare occasions when You decided to directly reveal a particularly important communication for the benefit of mankind — the burning bush, the annunciation, the Bhagavad-Gita, the giving of the Koran or the Golden Tablets, for instance — why did You neglect to mention the value of π , the reason for the seasonal agricultural cycle, or the biological basis of inheritance?"

What would you have written on your gravestone?

His self-love was, on occasion, required. ■

Double trouble for neurons

Mark E. Fortini

Mutations in the presenilin protein are associated with Alzheimer's disease. It now seems that mutant presenilin could wreak havoc on neuronal functions by triggering the activation of certain genes.

Working in Frankfurt, Germany, in 1901, Alois Alzheimer encountered a middle-aged patient named Auguste D. suffering from confusion and memory loss¹. Alzheimer's curiosity about her symptoms would lead to the recognition of the neurological syndrome that now bears his name. More than a century later, Alzheimer's disease is fixed in the public's consciousness as a devastating senile dementia that stubbornly resists effective medical treatment. But stunning advances are now being made in unravelling the molecular causes of this disease. Writing in *Cell*, Marambaud *et al.*² present evidence for a new molecular mechanism that is connected with Alzheimer's disease. They suggest that N-cadherin — a cell-surface protein that helps neurons adhere to one another — might relay signals associated with neuronal survival and 'plasticity', processes that underlie learning and memory. This signalling depends on the protein presenilin, mutations in which account for some cases of Alzheimer's disease.

Presenilin has been implicated in the 'amyloid theory' of how Alzheimer's disease develops. This view holds that a cell-surface protein — amyloid precursor protein (APP) — sheds sticky bits of itself from neurons and forms toxic aggregates known as amyloid plaques³. At the centre of this process is a large protein complex termed γ -secretase, which is composed of presenilin and a few other proteins⁴. Presenilin is an aspartyl protease, a protein that cleaves other proteins and relies on catalytically active aspartyl amino acids to do so. Its protein chain spans the membrane many times, and it cleaves its substrate proteins within the hydrophobic lipid bilayer of membranes — an environment generally thought to be unfavourable for breaking peptide bonds⁵. Just this sort of cleavage is behind the production of the sticky amyloid peptides seen in Alzheimer's disease, which are secreted from neurons after γ -secretase and other proteases act together to snip the small peptides off APP (Fig. 1a).

Marambaud *et al.*² now raise the provocative idea that the secreted amyloid peptides might have an intracellular accomplice. The authors reveal that N-cadherin is also cleaved by γ -secretase. At first, this finding seems unsurprising, as over the past

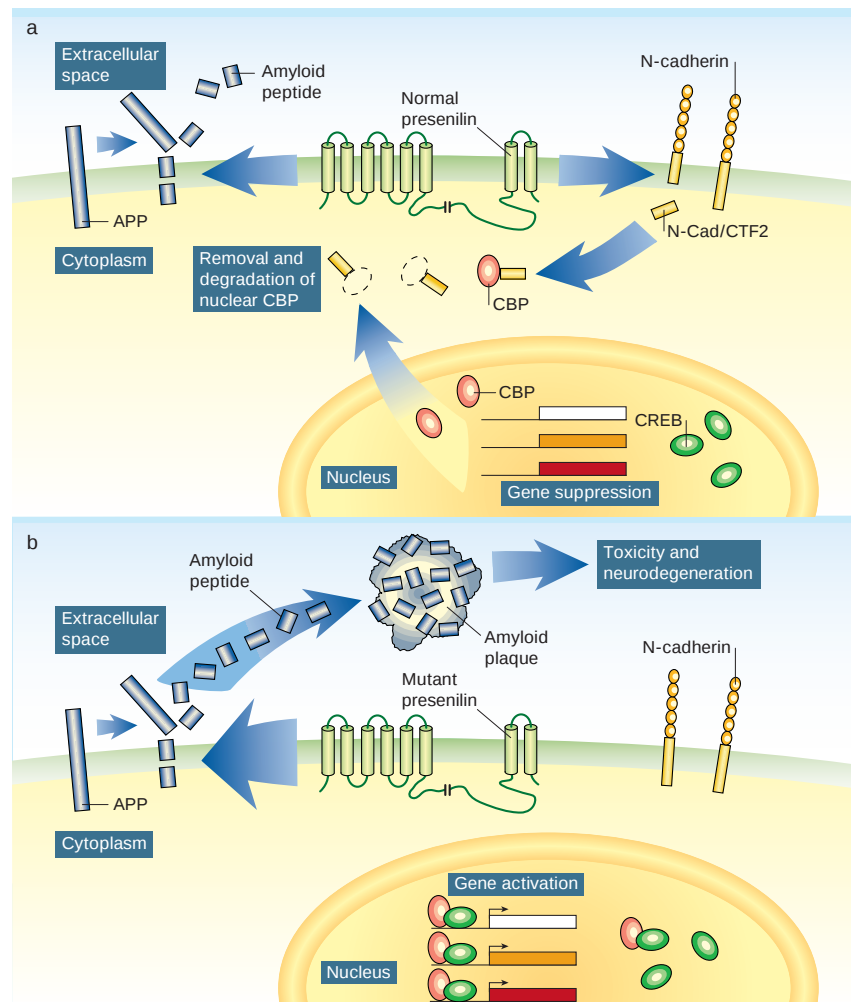


Figure 1 Two functions of presenilin. a, The presenilin protein, embedded in the outer membrane of neurons, is part of the γ -secretase complex, which snips fragments off amyloid precursor protein (APP), releasing amyloid peptide into the extracellular space. Marambaud *et al.*² show that presenilin also snips off the intracellular portion of the membrane protein N-cadherin. The released fragment, N-Cad/CTF2, binds to the protein CBP and targets it for destruction. In the absence of CBP, the nuclear protein CREB cannot activate genes. b, Mutated forms of presenilin that are associated with Alzheimer's disease are thought to accelerate the production of toxic amyloid plaques. These presenilin mutants also fail to cleave N-cadherin, increasing levels of CBP, and increasing the activity of genes regulated by CREB². So, as well as provoking amyloid-related neurotoxicity, presenilin mutations might cause Alzheimer's disease by inappropriately activating genes associated with neuronal function and memory formation.

few years a plethora of new γ -secretase substrates has been identified, including cell-surface receptors that regulate several major developmental and cancer-related pathways⁴. Indeed, it is established that γ -secretase cleaves the related adhesion

protein E-cadherin^{6,7}, thereby dismantling the adhesive contacts between cells.

But the cleavage of N-cadherin gives a new dimension to the presenilin story. In experiments with cultured mouse neurons, Marambaud *et al.* showed that N-cadherin

cleavage requires presenilin and is sensitive to neuronal stimulation — cleavage is induced by treatments that cause ion influx into neurons and trigger the release of neurotransmitters, the chemical messengers that propagate electrical signals from one neuron to the next. The result of N-cadherin cleavage is also intriguing. Part of this cell-surface protein protrudes through the cell membrane into the cytoplasm (the intracellular space), and cleavage releases this biologically active intracellular domain. Marambaud *et al.* found that this portion, N-Cad/CTF2, represses the activity of another protein termed CBP, a 'transcriptional coactivator' that helps another protein, CREB, do its job of turning on sets of target genes.

This is an exciting finding, as CBP and CREB are widely expressed proteins that directly or indirectly regulate many genes during normal cell growth and disease. CBP has also been linked to human neurodegenerative disorders and mental retardation⁸. And CREB-regulated gene expression has an enduring, albeit elusive, connection to the molecular basis of learning and memory⁹.

How does N-Cad/CTF2 influence CBP and CREB, and what are the consequences for neuronal gene regulation? The authors² present a series of experiments indicating that N-Cad/CTF2 binds to CBP in the cytoplasm, marking it for degradation and thus reducing the total cellular pool of CBP. An odd feature of this mechanism is that CBP is normally detected only in the nucleus, so N-Cad/CTF2 must either target newly synthesized CBP, or it must somehow 'pull' nuclear CBP into the cytoplasm to destroy it. CREB is not directly affected by N-Cad/CTF2, but without its partner, CBP, it is a less potent gene activator.

The notion that γ -secretase cleaves N-cadherin, generating a cytoplasmic protein fragment that influences the activity of many genes, including those potentially involved in memory formation, has far-reaching implications. At the very least, these findings will reinvigorate the debate about the relative importance of extracellular amyloid-related processes and intracellular signalling and gene regulation in Alzheimer's disease. Marambaud *et al.* report that several different presenilin mutations that are associated with the disease impart two characteristics to cells — the failure to produce N-Cad/CTF2 and the concomitant inability to suppress CREB-mediated gene activation (Fig. 1b).

So according to the new model, the presenilin mutations provoke a gain-of-activity change in numerous genes, which might possibly impair memory and other aspects of neuronal function, without necessarily killing neurons through toxic amyloid overload. Tantalizing evidence in support of this idea is that certain rare presenilin mutations

are linked to a form of dementia that is not associated with pronounced amyloid-plaque formation¹⁰. Furthermore, CBP- and CREB-regulated gene expression is often perturbed in other human neurodegenerative diseases¹¹.

Many details of this model, such as which CREB-responsive genes are most susceptible to regulation by N-Cad/CTF2 and whether this protein also affects other gene-activation pathways that do not use CREB, have yet to be addressed. Whether N-Cad/CTF2 induces CBP destruction in animals also awaits confirmation, as Marambaud *et al.* deduced this effect from studies in cultured cells. But as a promising first step in this direction, the authors examined neurons from genetically engineered mouse embryos that harbour an Alzheimer's disease-associated presenilin mutation. These mouse cells contained reduced amounts of N-Cad/CTF2 and increased levels of both nuclear CBP and the protein c-Fos (which is encoded by a gene that is activated by CREB) (Fig. 1b).

The new findings² also raise the question of whether presenilin mutations contribute to Alzheimer's disease by influencing gene-regulatory pathways other than those affected by N-cadherin cleavage. Indeed, several other γ -secretase substrates (including APP,

the Notch receptor protein, and the Deleted in Colon Cancer protein) are known to modulate various aspects of neuronal form and function. Could the cumulative effect of interfering with several such pathways trigger early neuron damage, before detectable amyloid plaques form and kill neurons? If so, are these pathways relevant to sporadic Alzheimer's disease, which is not associated with presenilin mutations? Future research might provide the answers to these questions. But as the new study illustrates, after almost a century of investigation, Alzheimer's disease has been slow to reveal its secrets. ■

Mark E. Fortini is in the Laboratory of Protein Dynamics and Signalling, National Cancer Institute, 1050 Boyles Street, Frederick, Maryland 21702, USA.

e-mail: fortini@ncifcrf.gov

1. O'Brien, C. *Science* **273**, 28 (1996).
2. Marambaud, P. *et al.* *Cell* **114**, 635–645 (2003).
3. Selkoe, D. J. *Trends Cell Biol.* **8**, 447–453 (1998).
4. Sisodia, S. S. & St George-Hyslop, P. H. *Nature Rev. Neurosci.* **3**, 281–290 (2002).
5. Wolfe, M. S. & Haass, C. *J. Biol. Chem.* **276**, 5413–5416 (2001).
6. Marambaud, P. *et al.* *EMBO J.* **21**, 1948–1956 (2002).
7. Baki, L. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 2381–2386 (2001).
8. Goodman, R. H. & Smolik, S. *Genes Dev.* **14**, 1553–1577 (2000).
9. Kandel, E. R. *Science* **294**, 1030–1038 (2001).
10. Amtul, Z. *et al.* *Neurobiol. Dis.* **9**, 269–273 (2002).
11. Freiman, R. N. & Tjian, R. *Science* **296**, 2149–2150 (2002).

Cosmology

The shape of the Universe

George F. R. Ellis

An analysis of astronomical data suggests not only that the Universe is finite, but also that it has a specific, rather rigid topology. If confirmed, this is a major discovery about the nature of the Universe.

What shape is space? On page 593 of this issue, Luminet *et al.*¹ suggest that the topology of the Universe may be a 'Poincaré dodecahedral space' — as illustrated on this week's cover. And this is no idle abstraction: Luminet *et al.* show that this topology, unlike many others, is supported by data from NASA's Wilkinson Microwave Anisotropy Probe (WMAP), published earlier this year².

In thinking about the large-scale shape of the Universe, three interlinked questions must be confronted. First, what is its spatial curvature? There are three possible answers. Three-dimensional sections of space-time may be 'flat' — in such space sections, parallel lines stay the same distance apart and never meet (as in Euclidean space). Or the space sections may be 'negatively curved', such that parallel lines diverge from one another and never meet (the three-dimensional analogue of a Lobachevsky space). Finally, they may be 'positively curved', such that parallel lines converge and eventually intersect (the three-dimensional analogue of

the surface of a sphere). The particular case that exists depends on how well the amount of matter in the Universe, coupled with the driving force of dark energy, balances the Universe's kinetic energy of expansion. This is usually expressed in terms of the normalized density parameter Ω_0 , which is unity for flat space sections; for positive spatial curvature, Ω_0 is greater than one.

The second question is whether the Universe is 'open' or 'closed' — that is, is it spatially infinite, containing an infinite amount of matter, or is it spatially finite, containing a finite amount of matter? Positively curved space sections are necessarily closed, but the converse does not necessarily follow: both flat and negatively curved space sections can be finite if their connectivity is more complicated than in Euclidean space, meaning that their topology is quite unusual^{3,4} (for example, in a flat toroidal space, as you exit right you enter left, and space is finite). So the third issue is, what is the large-scale topology of the Universe?

It is worth noting that none of these

features is determined by the Einstein gravitational field equations, which are differential equations that govern local, rather than global, properties of space-time⁵. Topology and curvature seem to be fixed by the initial conditions at the start of the Universe that have since determined its dynamical evolution. To investigate the topology and curvature of the Universe, we must use astronomical observations; from observed values of the energy densities and the expansion rate in the Universe, the curvature can be deduced using Einstein's field equations.

If the Universe is closed, and has a small enough diameter, we may be able to see right round it because photons can traverse the whole Universe — Luminet *et al.*¹ illustrate this point well with the image of an insect crawling around the surface of a cylinder (Fig. 2 on page 593). If this is so, we may be able to identify multiple images of the same structure but in different directions in the sky^{4,6}, or see an effect on the statistics of clustering of galaxies⁴. Furthermore, the existence of such a 'small universe' should be detectable through its influence on that Rosetta Stone of present-day cosmology, the cosmic microwave background (CMB) radiation^{7,8}.

The CMB is the relic radiation of the Big Bang, and the very slight changes in its temperature across the sky record the density fluctuations that existed at a certain point in the early history of the Universe. Data on these anisotropies^{2,9} from WMAP and other sources suggest that the density parameter Ω_0 has the value 1.02 ± 0.02 . This is compatible with flat space sections. Further data might move the value of Ω_0 towards or even below unity. But taking the present data at their face value, the conclusion is that the Universe is spatially closed and hence has finite space sections. This tentatively answers two of the three questions mentioned above.

But there is a further intriguing feature in the data. The so-called power spectrum of anisotropies (Fig. 1) shows a distinctive set of peaks when the anisotropy is compared between regions of sky separated by small angles. But on large angular scales (for regions typically more than 60° apart), there is a strange loss of power that does not fit with the expectations of standard cosmological models (in particular, there is less power in the quadrupole than expected). Some have proposed that this can be attributed to as yet undiscovered laws of physics at work in the early Universe¹⁰.

Luminet *et al.*¹, however, suggest it is because we live in a universe with positively curved space sections and non-standard topology. The smaller-than-usual diameter of the Universe means that there is a maximum length scale that fits into it — and consequently there is a loss of power in the CMB spectrum on scales that are larger than this maximum⁷. The authors propose

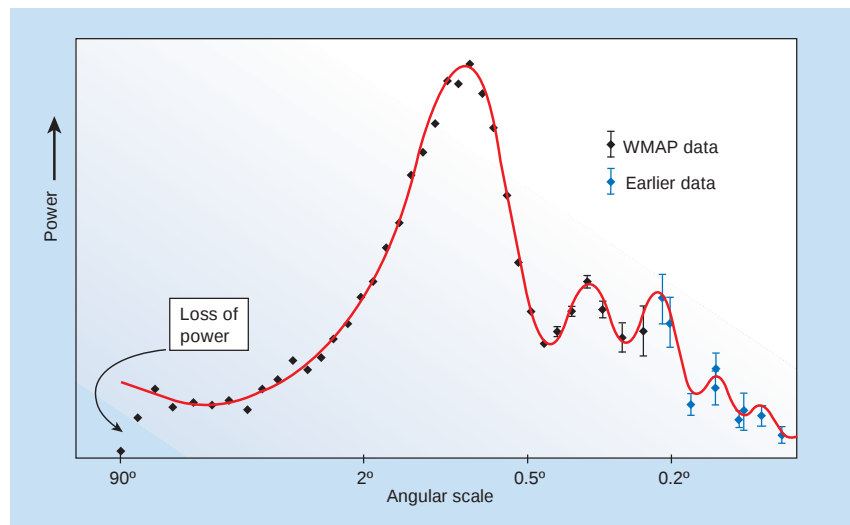


Figure 1 Power spectrum of the cosmic microwave background (CMB) radiation. Data² from the Wilkinson Microwave Anisotropy Probe (WMAP) have extended the accuracy of the spectrum far beyond what was known from earlier measurements. This plot reflects the small differences in the temperature of the CMB across the sky. There are a series of peaks in the spectrum at small angular separations, but at large scales that structure disappears, and the data were expected to follow the 'Sachs–Wolfe plateau'. The WMAP measurements, however, fall below this plateau at the largest angular scales. Standard cosmological models cannot explain this, but Luminet and colleagues' topological model¹ for a finite universe can.

that the spatial sections of the Universe are dodecahedral sections of a space of positive curvature, fitted together to make finite three-dimensional spaces. This topology accounts for the WMAP data better than do standard models¹.

Can this proposal be confirmed? Yes indeed. First, Luminet and colleagues' model suggests that $\Omega_0 = 1.013$, and future observations will produce data that should pin down the value of Ω_0 to this level of accuracy (the current value from WMAP is accurate only to 2%). Second, a remarkable paper by Cornish *et al.*¹¹ showed that, however complex the spatial topology, in a small universe there will be circles of identical temperature fluctuations in the CMB sky that could be identified from data on the CMB anisotropies. Luminet *et al.* determine what those circles would be in their model. Future analyses of WMAP data — and of data from its successor, the Planck satellite, to be launched in 2007 — should be able to verify whether the circles are there or not.

In many models of the Universe, it is assumed that spatial homogeneity extends outside our visual horizon for ever. However, in the case of chaotic inflation^{12,13} — a variant of the inflationary model for an exponential expansion that occurred in the wake of the Big Bang — the Universe is very inhomogeneous on scales much larger than we can observe, and we are in one expanding bubble in the middle of innumerable other, similar ones. But if Luminet *et al.* are correct, chaotic inflation is ruled out: there is only one expanding universe bubble, and we can see almost all the way round it.

In 1917, Einstein¹⁴ proposed that spatially closed universes are advantageous because that would remove the problem of boundary conditions at infinity¹⁵. A small universe in which we have seen most of what exists is even more advantageous⁶; indeed, strictly speaking, they are the only universes in which we can predict the astronomical future — the return of Halley's comet, for example — because only in them do we have access to all the data needed to make such predictions. The WMAP data, as interpreted by Luminet *et al.*¹, suggest that we might indeed live in such a small closed universe. ■

George F. R. Ellis is in the Department of Mathematics, University of Cape Town, Rondebosch 7700, Cape Town, South Africa.
e-mail: ellis@maths.uct.ac.za

- Luminet, J.-P., Weeks, J. R., Riazuelo, A., Lehoucq, R. & Uzan, J.-P. *Nature* **425**, 593–595 (2003).
- Bennett, C. L. *et al.* *Astrophys. J. Suppl. Ser.* **148**, 1–27 (2003).
- Ellis, G. F. R. *Gen. Rel. Grav.* **2**, 7–21 (1971).
- Lachieze-Rey, M. & Luminet, J. P. *Phys. Rep.* **254**, 135–214 (1995).
- Friedmann, A. Z. *Phys. Z.* **21**, 326–332 (1924); transl. *Gen. Rel. Grav.* **31**, 2001–2008 (1999).
- Ellis, G. F. R. & Schreiber, W. *Phys. Lett. A* **115**, 97–107 (1986).
- Riazuelo, A., Uzan, J.-P., Lehoucq, R. & Weeks, J. Preprint at <http://arXiv.org/astro-ph/0212223> (2002).
- Uzan, J.-P., Riazuelo, A., Lehoucq, R. & Weeks, J. Preprint at <http://arXiv.org/astro-ph/0303580> (2003).
- Spergel, D. N. *et al.* *Astrophys. J. Suppl. Ser.* **148**, 175–194 (2003).
- Tsujiikawa, S., Maartens, R. & Brandenberger, R. H. Preprint at <http://arXiv.org/astro-ph/0308169> (2003).
- Cornish, N., Spergel, D. & Starkman, G. *Class. Quant. Grav.* **15**, 2657–2670 (1998).
- Linde, A., Linde, D. & Mezhlumian, A. *Phys. Rev. D* **49**, 1783–1826 (1994).
- Guth, A. H. *Phys. Rep.* **333**, 555–574 (2000).
- Einstein, A. *Preuss. Akad. Wiss. Berlin Sitzber.* 142–152 (1917); transl. in *The Principle of Relativity* 177–188 (Dover, Mineola, New York, 1923).
- Wheeler, J. A. *Einstein's Vision* (Springer, Berlin, 1968).

Clocking the birth of neurons

William A. Harris

Different types of neurons are born in a conserved, sequential order during development. The molecular cogs in the clock-like mechanism driving this process are now being revealed.

Millions of nerve cells, of many different varieties, are generated in the central nervous system during development. The cells that give rise to this plethora of neurons, known as neuroblasts, follow certain temporal guidelines: they generate particular types of neurons early in development, and other types later. For example, the neurons in the lower layers of the brain's cerebral cortex are born early, and those in the top layers are born later. This sequence is thought to have something to do with a progressive loss of a neuroblast's intrinsic capability, or 'competence', because when, for instance, older neuroblasts are transplanted into a young cortex in ferret embryos, they cannot make lower-layer neurons¹. Thanks to the work of Pearson and Doe² (page 624 of this issue), we now have some idea about the molecular basis of this change, at least for one particular neuroblast in fruit-flies. The authors show that the transient expression of a single protein, called Hunchback, regulates the progressive loss of competence.

Most of the neuroblasts in the fruit-fly *Drosophila melanogaster* have been

identified and assigned numbers, and it is possible to follow them and their progeny individually through time-lapse confocal microscopy, using flies that have been engineered to express fluorescent proteins indelibly in specific cell lineages. Let us look at neuroblast NB7-1. Microscopy has shown that this neuroblast generates five successive cells known as ganglion mother cells (GMCs): we call them GMC1 to GMC5. Each GMC then produces two post-mitotic (non-dividing) neurons. U1 is a motor neuron that innervates dorsal muscles. It is born from GMC1, the first progeny of NB7-1. Meanwhile, U5 — which innervates other muscles — is born from GMC5, the last progeny of NB7-1 (Fig. 1a).

What is behind this maturation of NB7-1? Two years ago, Doe and colleagues³ showed that, as this cell churns out its neuronal issue, it changes its molecular nature like a magic ball that changes colour. When NB7-1 is young and giving rise to GMC1, it expresses Hunchback (Hb), a protein that controls gene activity. When it is old and giving rise to GMC5, it expresses another gene-transcription factor, called Castor. In between, other transcription factors are successively

expressed. And each GMC matches the 'colour' of its mother at the time of its birth by the transcription factors that it expresses. So GMC1 expresses Hb, as its mother did when it was born, whereas GMC5 expresses Castor. Hb expression is essential for producing early neurons such as U1; in Hb mutants there are no U1s. Similar results show that the successive transcription factors are essential for producing the successive types of neuron.

But how does a neuroblast such as NB7-1 change its colours as a function of time — and when it's out of the blue and into the black, can it ever come back? In other words, what happens if Hb is experimentally re-expressed in an old NB7-1 that has turned this protein off? This is the problem that Pearson and Doe² have tackled.

The authors used genetically engineered flies in which they altered the timing of Hb expression. They found that if they allowed an NB7-1 to age so that it no longer naturally expressed Hb, and then turned this protein on again, the old NB7-1 had completely lost the competence to produce U1s (Fig. 1b). But 'middle-aged' NB7-1s, in which there was still a little Hb expression, only partially lost this competence, and an extra burst of Hb in these neuroblasts led to extra U1s.

So the loss of competence to produce early cell types correlates with the normal downregulation of Hb. If Hb levels are never allowed to fall — as in a genetically engineered fly in which Hb is expressed continuously throughout the lifetime of NB7-1 — the neuroblast seems to stay forever young and produces U1s from every GMC it generates. Pearson and Doe also found that Hb needs to be in the dividing neuroblast or GMC to maintain the youthful state. It was no use adding extra Hb to a maturing U2: by the time this neuron is post-mitotic, it has already lost the competence to respond to Hb, and cannot now become a U1.

This work gives a glimpse into the molecular nature of an intrinsic clock-like mechanism that ticks through developmental time in a neuroblast, leading to a progression of distinct neuronal fates. It is known that transient expression of Hb in flies, or of the related Ikaros protein in mammals, can silence gene expression by interacting with a complex in the cell nucleus that removes acetyl groups from histones^{4,5} (histones being proteins that help to package DNA). So Pearson and Doe² propose that Hb might, in this way, be silencing 'late-fate' genes such as Castor. The speculation is that when Hb levels go down, the late-fate genes become unsilenced — and once they do, they can no longer be turned off. Clearly, we need to know more about this mechanism. Does something similar happen at each tick of the neuroblast clock? How do the

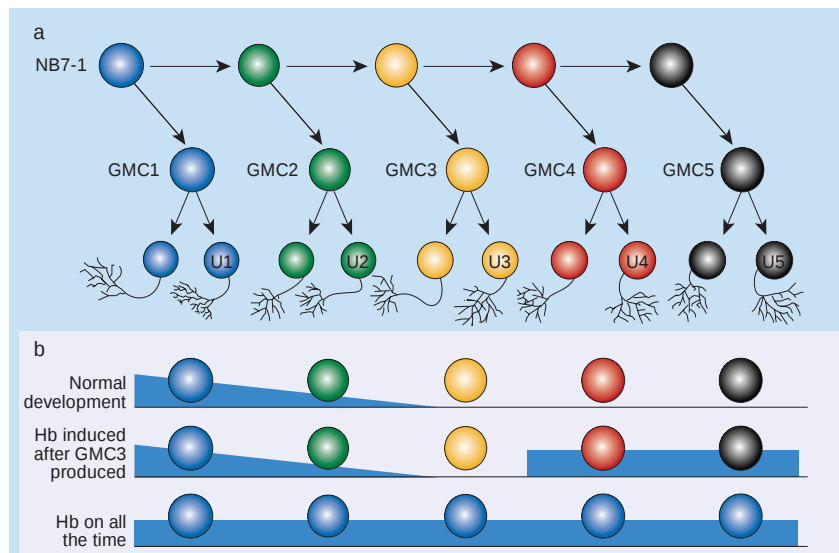


Figure 1 The ageing neuroblast. **a**, As neuroblast NB7-1 ages, it gives rise to a defined sequence of ganglion mother cells (GMCs), which in turn produce a specific sequence of U-type motor neurons, among other nerve cells. **b**, The molecular controls behind the ageing of NB7-1. Top, levels of the Hunchback protein (Hb) gradually fall during normal development. Centre, Pearson and Doe² show that if Hb is experimentally switched back on just after the birth of the Hb-negative GMC3, the neuroblast continues to age as normal. Bottom, the authors also find that if NB7-1 is engineered to express Hb continuously, it remains young and able to produce early neurons.

ticks correspond to the rounds of the cell-division cycle? And is something similar at work in other ageing neuroblasts, and in other animal species?

William A. Harris is in the Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK.

e-mail: harris@mole.bio.cam.ac.uk

1. McConnell, S. K. *Neuron* **15**, 761–768 (1995).
2. Pearson, B. J. & Doe, C. Q. *Nature* **425**, 624–628 (2003).
3. Ishiki, T., Pearson, B., Holbrook, S. & Doe, C. Q. *Cell* **106**, 511–521 (2001).
4. Sabbatini, P. *et al.* *EMBO J.* **20**, 2812–2822 (2001).
5. Zhang, C. C. & Bienz, M. *Proc. Natl Acad. Sci. USA* **89**, 7511–7515 (1992).

Plant biology

Locks, keys and symbioses

Martin Parniske and J. Allan Downie

The association between legumes and nitrogen-fixing bacteria requires molecular recognition to allow bacterial entry into root hairs. The discovery of a novel type of plant receptor clarifies how this happens.

Legumes, such as pea, bean, trefoil and peanut, are agricultural wonders. They form symbioses with bacteria, known as rhizobia, which means that they can make their own nitrogen fertilizer by 'fixing' atmospheric nitrogen. The fixation process occurs in root nodules. But this is the end-point of a developmental programme, which starts with a molecular recognition system that enables root hairs to promote the entry of rhizobia while excluding a huge diversity of unwelcome intruders from the soil. In 1990, the structure of one component of the recognition system was published¹. That component, which comes in various types, is the Nod factor. It is produced by rhizobia and is the key that they need to open the root-hair lock and enter the plant.

Since 1990, identification of the corresponding plant receptor for Nod factors has been a central goal of research. Papers by Radutoiu *et al.*² and Madsen *et al.*³ (pages 585 and 637 of this issue), and a report in *Science* by Limpens *et al.*⁴, now describe a new class of receptor kinases that could constitute the lock for the Nod-factor key. Kinases are enzymes that add phosphate groups to other proteins: they act as molecular switches, which turn enzymes or signalling pathways on or off.

In legumes, the root hairs exhibit an astonishing developmental switch once molecular recognition occurs between rhizobia and a plant, and the entrance mechanism is triggered. The tip of a root hair stops growing in a straight line and the cell curls back on itself, trapping bacteria within a pocket. The bacteria grow in the pocket and are taken up into a plant-made intracellular 'infection thread' (Fig. 1). Overall, the root-hair response involves a complex redirection of cellular development, requiring orchestration by positional and chemical cues.

Radutoiu *et al.*² worked with mutants of the legume *Lotus japonicus*. They found that plants with mutations in two genes, *NFR1* or *NFR5*, do not respond to Nod factor and are not infected by rhizobia. Nod factors have a

backbone of several *N*-acetylglucosamine molecules, which are 'decorated' with various other chemical groups depending on the Nod factor concerned. Radutoiu *et al.* demonstrate that *NFR1* and *NFR5* proteins carry two or three extracellular LysM motifs. The significance of this is that the LysM motif occurs in a variety of enzymes that bind to polymers containing *N*-acetylglucosamine^{5,6}. So it seems that the *NFR* proteins bind Nod factors through their extracellular LysM domains, and that binding elicits signalling via the intracellular kinase domain. The LysM receptors probably act in concert with

other players such as SYMRK^{7,8}, another receptor kinase, but one that lacks LysM motifs in its extracellular domain.

Radutoiu *et al.* show that the earliest detectable root-hair responses to Nod factor, typically observed within minutes, are completely abolished in *nfr5* mutants and modified in *nfr1* mutants, but are retained in a *symrk* mutant. Given this, it looks as if *NFR5* and *NFR1* act sequentially and individually, followed by SYMRK. But it is also possible that different combinations of receptors form recognition complexes. This would resolve one puzzle — the *NFR5* kinase does not have an 'activation loop', a usual feature of these enzymes; therefore, because it is predicted to lack kinase activity, *NFR5* probably relies on an interaction with another protein to initiate signalling.

The *NFR1* and *NFR5* proteins are clearly required during the earliest stages of the rhizobium–plant interaction (Fig. 1a,b). But other results indicate that LysM receptors are also involved in later events. Rhizobia continue to produce Nod factor within infection threads, and Madsen *et al.*³ show that the pea version of *NFR5*, *SYM10*, is expressed at a higher level in tissue that contains growing infection threads. But we cannot yet tell whether *NFR5*/*SYM10* is required for sustained infection-thread growth because of the block on thread initiation in *nfr5/sym10* mutants.

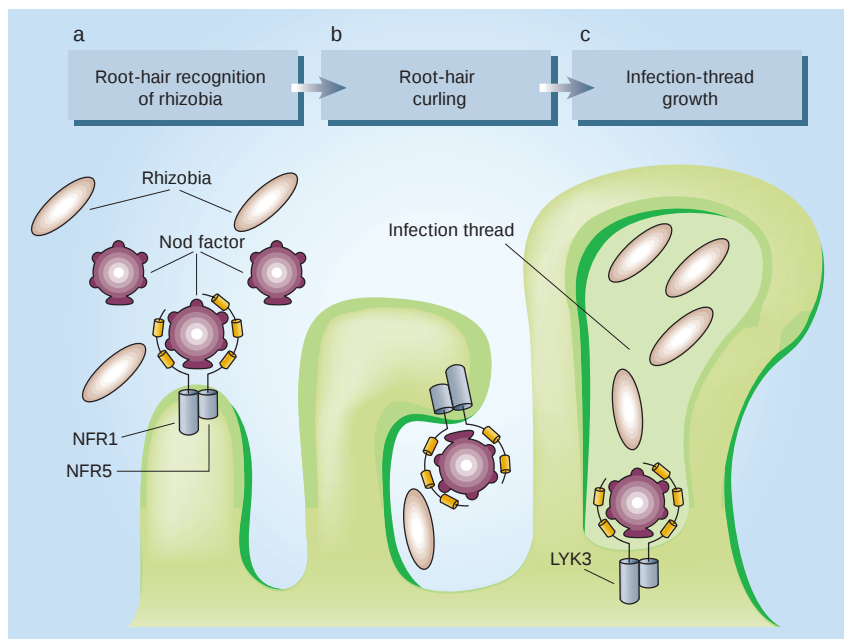


Figure 1 Unlocking the root-hair door. a, b, Two receptors enable a plant to sense the Nod factor produced by the 'right' symbiotic rhizobia, and then allow root-hair curling for the rhizobia to enter. Work with *Lotus japonicus* shows that the receptors, *NFR1* and *NFR5*, act at the earliest stage in events^{2,3}. c, The probable *Medicago truncatula* variant of *NFR1*, *LYK3*, appears to be required to maintain infection-thread growth⁴; the expression pattern³ of the pea version of *NFR5*, *SYM10*, suggests that this protein might also be required at this stage. But it remains to be seen if *NFR1*, *NFR5* and/or *LYK3* interact as speculated here. Overall, it looks as if there could be a continuous requirement for *NFR1* and *NFR5*, or their counterparts in other legume species, throughout root-hair curling and infection-thread formation.

In their work, Limpens *et al.*⁴ started with a specific line of pea that has an unusual pattern of Nod-factor recognition. The root hairs curl and entrap rhizobia. But the process stops after infection-thread initiation unless the rhizobia produce a Nod factor that has specific decorations^{9,10}. Limpens *et al.* mapped the overall genetic locus, *SYM2*, responsible. In another species, *Medicago truncatula*, which has a smaller genome, they narrowed down the *SYM2* target region to a stretch of DNA about 300 kilobases in length. Surprisingly, it contained many excellent candidates for involvement in Nod-factor perception — among them seven receptors potentially involved in disease resistance, and seven receptor kinases with extracellular LysM domains. Each candidate gene was silenced using an RNA interference strategy. Only with one, called *LYK3*, did a symbiotic defect result: when *LYK3* expression was reduced, Limpens *et al.* found that although infection-thread growth was initiated, it subsequently ceased. This effect depended on the structure of the Nod factor produced by the bacteria.

The authors' interpretation is that there is a second lock-and-key mechanism in the developing infection thread, and that the second lock is encoded by *LYK3* (Fig. 1c). Because of similarity in sequence, and a similar genomic environment, this gene is likely to be the *Medicago* version of the *Lotus* *NFR1*. Curiously, however, *Medicago* plants with silenced *LYK3* and *Lotus nfr1* mutants differ substantially. One possible explanation stems from Limpens and colleagues' observation that silencing of *LYK3* in their system was incomplete. The residual *LYK3* expression could be sufficient to establish the first lock, but not the second. Once pea variants of *LYK3* have been identified, it will be possible to find out whether they contribute to the unusual pattern of Nod-factor recognition of the pea line that initially sparked the interest in the *SYM2* genetic locus.

There is also an evolutionary angle to the new findings. Chemical decoration of Nod factors varies depending on the rhizobial strain, and it determines the specificity between a host plant and its bacterial symbiont¹¹. The genes encoding Nod-factor receptors appear to evolve rapidly, because very closely related plants can prefer different Nod-factor structures. *LYK3* (ref. 4) and *NFR1* (J. Stougaard, personal communication) reside in clusters of several similar genes, in tandem array, which may be the clue to how the corresponding sequence diversification is achieved. By facilitating recombination between diverged gene copies within the array, such an arrangement is suitable for generating new sequence variants, and new recognition specificities, relatively quickly¹².

There is much to learn. In particular, it remains to be seen exactly how the different LysM kinases integrate into a signalling

pathway (or network) to guide root-hair curling, and initiate and sustain infection-thread growth. Another major task is to find out whether, or how, Nod factors and LysM receptor kinases physically interact. ■

Martin Parniske is in the Sainsbury Laboratory, John Innes Centre, and J. Allan Downie is in the Department of Molecular Microbiology, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK. e-mails: martin.parniske@sainsbury-laboratory.ac.uk allan.downie@bbsrc.ac.uk

1. Lerouge, P. *et al.* *Nature* **344**, 781–784 (1990).

2. Radutoiu, S. *et al.* *Nature* **425**, 585–592 (2003).

3. Madsen, E. B. *et al.* *Nature* **425**, 637–638 (2003).

4. Limpens, E. *et al.* *Science* published online 28 August 2003 (doi:10.1126/science.1090074).

5. Amon, P., Haas, E. & Sumper, M. *Plant Cell* **10**, 781–789 (1998).

6. Bateman, A. & Bycroft, M. *J. Mol. Biol.* **299**, 1113–1119 (2000).

7. Stracke, S. *et al.* *Nature* **417**, 959–962 (2002).

8. Endre, G. *et al.* *Nature* **417**, 962–966 (2002).

9. Firmin, J. L., Wilson, K. E., Carlson, R. W., Davies, A. E. &

Downie, J. A. *Mol. Microbiol.* **10**, 351–360 (1993).

10. Guerts, R. *et al.* *Plant Physiol.* **115**, 351–359 (1997).

11. Perret, X., Staehelin, C. & Broughton, W. J. *Microbiol. Mol. Biol. Rev.* **64**, 180–201 (2000).

12. Parniske, M. *et al.* *Cell* **91**, 821–832 (1997).

Nuclear physics

It's a knockout

David Warner

In collisions between nuclei, a proton or neutron might be knocked out of one nucleus. Now, two-proton knockout has been demonstrated, opening a new route to the creation of neutron-rich systems for study.

Most of our knowledge about the quantum nature of the atomic nucleus comes from the study of nuclear reactions in which an energetic beam of one nuclear species is directed onto a target made of another. Writing in *Physical Review Letters*, Bazin and colleagues¹ have now demonstrated that it is possible to knock two protons simultaneously out of a nucleus in the high-energy beam, while leaving the remaining nucleons (neutrons and protons) largely undisturbed. This observation of a so-called 'direct' reaction process offers a valuable tool to probe some of the most exotic regions of the nuclear chart.

When two nuclei react in a collision, there is a plethora of possible outcomes, because there are many different ways in which the interaction can take place. One class of reaction², referred to as 'direct', has proved crucial in elucidating the motion of the individual nucleons inside a nucleus. In a high-energy

direct reaction, the projectile undergoes a peripheral, grazing collision with the target nucleus, and only the surfaces interact. This type of reaction occurs quickly, compared with, say, a 'compound' reaction in which the two nuclei collide violently and the energy is shared out among all of the constituents of the combined system. In the direct reaction, there is then a simple relationship between the physical descriptions, known as wavefunctions, of the incoming and outgoing systems; in a knockout reaction, for example, they differ only through the addition or removal of one or two nucleons. The reaction thus creates a new nucleus and simultaneously provides detailed information about the quantum states that its nucleons occupy.

Until recently, the chief constraint on experiments exploring the nuclear many-body system was that, in any reaction chosen for study, both beam and target had to be stable, naturally occurring species. But the

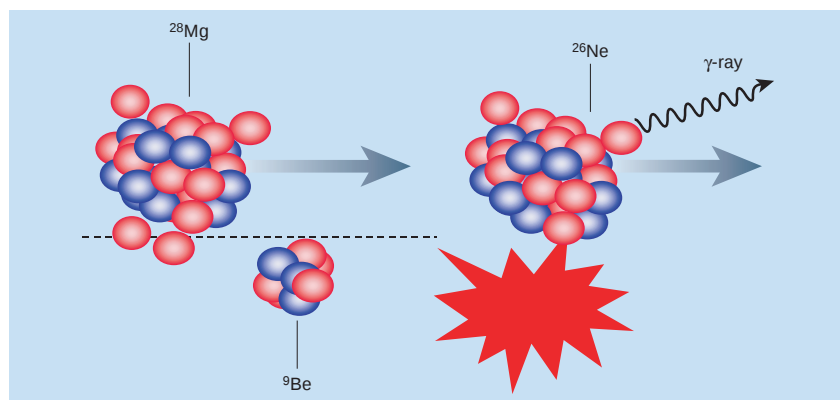


Figure 1 Two-proton knockout. Bazin *et al.*¹ have demonstrated that the incidence of a beam of unstable magnesium nuclei (^{28}Mg) on a target of stable beryllium nuclei (^9Be) produces a residue of neon (^{26}Ne) nuclei. During the reaction, two of the four loosely bound, outermost protons in the beam nucleus are removed, leaving the ^{26}Ne nucleus, which then emits γ -radiation. The velocity of the residue is the same as that of the beam particles.

advance of accelerator technology now means that it is possible to create high-energy beams of unstable nuclei. There is potentially an enormous range of radioactive nuclear beams available to induce nuclear reactions, to create nuclei with increasingly exotic combinations of neutrons and protons. This is already beginning to transform the field of nuclear physics, enabling physicists to address many outstanding questions.

Bazin's team¹ — working at one of the new generation of radioactive beam facilities, at the National Superconducting Cyclotron Laboratory in Michigan — directed a beam of the unstable magnesium isotope ²⁸Mg onto a natural beryllium target (⁹Be). The ²⁸Mg nucleus contains four more neutrons than the most abundant stable isotope of magnesium; the knockout of two protons creates a neon nucleus (²⁶Ne) that has an even larger excess of neutrons (Fig. 1). This is a crucial feature of the reaction — through this route, experimentalists should be able to access the most neutron-rich nuclei and probe their quantum structure. In heavier stable nuclei, there is a steadily increasing excess of neutrons over protons, to counteract the mutual repulsion between protons due to their electric charge. But artificially creating such heavy nuclei simply by combining a pair of light stable nuclei invariably produces a compound system that has a deficiency of neutrons. Hence neutron-rich systems have been least studied — although they have also revealed the biggest surprises, such as the discovery of a class of light, neutron-rich nuclei in which one or two neutrons form a diffuse 'neutron halo' with a radius as much as three times that of the core³.

The team confirmed the simple direct nature of the reaction process in two ways. The nuclear shell model tells us that the first eight protons of the twelve in a magnesium nucleus form an inert, tightly bound, inner core. Thus the reaction, if it is direct, can be expected to involve the four outer (valence) protons only. Bazin *et al.*¹ calculated the basic cross-section, or probability, for a process involving any two uncorrelated valence protons and then multiplied it by six — the number of pairs that can be formed from the four valence protons. The result was already within 20% of the measured value for the cross-section.

They applied the same model to the measured distribution of velocities of the neon nuclei produced in the reaction. The observation that this distribution was narrow and centred on the beam velocity gave qualitative confirmation that a peripheral, direct process was involved. Assuming that the two protons could be treated as two independent (uncorrelated) particles provided a convincing quantitative reproduction of the data.

Bazin *et al.*¹ went further. By detecting coincident γ -rays, a technique that they had

pioneered in earlier work on one-nucleon knockout reactions^{4–7}, they were able to extract the relative probabilities for populating individual quantum states in the exotic ²⁶Ne nucleus produced. Reproducing the detailed distribution of these probabilities in their calculations required a sophisticated treatment of the relative motion of the two protons removed in the reaction. But a good description of the data is achieved by assuming that the protons are plucked from the projectile as a two-proton cluster, in which the protons have no angular momentum relative to each other.

This study shows in a straightforward and convincing way that the two-proton knockout reaction generated with neutron-rich

radioactive beams is a direct process that can probe the detailed structure of states in increasingly exotic nuclei. Bazin and colleagues' work heralds an important technique to be exploited in a new generation of experiments with radioactive nuclear beams. ■

David Warner is in the Surface and Nuclear Division, Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, UK.

e-mail: d.warner@dl.ac.uk

1. Bazin, D. *et al.* *Phys. Rev. Lett.* **91**, 012501 (2003).
2. Satchler, G. R. *Direct Nuclear Reactions* (Oxford Univ. Press, 1983).
3. Hansen, P. G. *Nature* **384**, 413–415 (1996).
4. Navin, A. *et al.* *Phys. Rev. Lett.* **81**, 5089–5092 (1998).
5. Tostevin, J. A. J. *Phys. G* **25**, 735–739 (1999).
6. Maddalena, V. *et al.* *Phys. Rev. C* **63**, 024613 (2001).
7. Enders, J. *et al.* *Phys. Rev. C* **65**, 034318 (2002).

Neuroscience

Re-recording human memories

Karim Nader

Two studies help reveal the dynamics of memory. New memories that weaken during the day can be strengthened by a period of sleep. And when memories are reactivated, they must be re-stored in order to persist.

Memory research is undergoing a transformation. No longer is memory thought to be a hard-wiring of information in the brain. Instead, it seems to be a process of storage and re-storage¹. Two papers in this issue take us closer to understanding the dynamic nature of human memory. On page 614, Fenn *et al.*² show that sleep can rescue memories that were lost during the day. And on page 616, Walker *et al.*³ demonstrate that stabilized memories can be re-stored when they are reactivated — and that those memories are lost if this process is interfered with.

When we learn something new, the memory passes through two states. The first is called short-term memory. In this state, memories are 'labile' — interfering with their processing in the brain changes how the memory is stored. Over the following several hours the memory is stabilized, or consolidated⁴, by a process that requires neurons to synthesize new RNA and proteins (cellular consolidation)⁵. These long-term, consolidated memories are usually thought of as being fixed in the brain. A good example of memory transition is remembering a new telephone number: if you are distracted immediately after learning a new number, you are likely to forget it. But if the same distraction is delayed by a day you will probably remember the number. Immediately after learning the number, the memory is labile and sensitive to disruption, whereas 24 hours later the memory is consolidated and resistant to the same disruption.

Sleep has been linked to memory consolidation⁶ — it can often enhance the memory of information acquired the previous day. So

sleep helps some memories 'mature' and also prunes out unimportant memories. But the benefits obtained from sleep are typically thought to be restricted to previously learned information. In other words, the sleep-enhanced memory of a particular sequence of movements (motor sequence) does not improve the performance of other motor sequences⁷.

Fenn *et al.*² demonstrate a new effect of sleep on memory. The authors suggest that, as well as enhancing memories, sleep can rescue memories that are weakened during the preceding day. Remarkably, they also show that this benefit is not restricted to the specific information acquired on that day — the rescued memories can be applied to improve performance in new situations.

The authors looked at the effect of sleep on the performance of a perceptual learning task. Human subjects were asked to identify a series of phonetically similar words produced by a synthetic speech machine. The subjects never heard the same word twice, so their improvement in identifying new words depended on their ability to apply what they had learned during a previous training period. In other words, their ability to generalize was tested.

After a single training period during the morning, the subjects showed a 21% increase in the accuracy with which they could identify new words. After 12 waking hours, that accuracy was reduced to 10%, but accuracy was not decreased if the 12-hour interval included a period of sleep. And, crucially, the decrease in accuracy seen after 12 waking hours could be recovered with a subsequent period of sleep. Additional

groups that were trained either in the morning or evening and then tested 24 hours later also showed an 18% increase in accuracy, indicating that the improvement was not due to general circadian effects on learning.

These results extend another finding — that power naps can reverse the detrimental effects of overtraining⁸ — in two ways. First, Fenn *et al.* show that sleep can rescue memories that have spontaneously deteriorated, and second, they demonstrate that memories involving generalization can be recovered. It will be interesting to see what other kinds of memories can be rescued by sleep and what phases of sleep cause this effect.

But how does sleep enhance memory? Cellular consolidation takes only a few hours. So according to theory, most memories should be fixed in the brain and insensitive to further manipulation before a period of sleep occurs. Walker *et al.*³ examined the consolidation profile of a motor memory by looking at interference between two related memories. Interference — when the acquisition of a second memory impairs the memory of the original task — is time dependent and is a classical approach to determining whether memories are labile or consolidated⁴.

The authors show that if a second finger-tapping exercise is learned immediately after a first exercise, the memory of the second sequence interferes with that of the first, and the performance of only the second sequence is enhanced by sleep. The memory of the second sequence is not somehow erasing the memory of the first, as the subjects performed the first exercise normally if they were tested immediately after the training for the second exercise. In contrast, if the second sequence is learned six hours after the first sequence, interference does not occur and the performance of both sequences is enhanced by sleep.

This finding demonstrates that the memory of the first sequence becomes consolidated within six hours. So how can a subsequent period of sleep enhance this already 'fixed' memory? The authors have previously suggested that there is a second time window for consolidation during sleep, but how can fixed memories become sensitive to further processing?

The answer probably includes the fact that reactivating a consolidated memory can return it to a labile state from which it needs to undergo cellular re-storage or reconsolidation in order to persist. For example, if auditory fear memories are reactivated in rats, protein synthesis is required for these memories to persist. If the fear memories are not reactivated, inhibiting protein synthesis has no effect on their persistence, demonstrating that reconsolidation is predicated on memory reactivation¹. So memory storage seems to occur more than once.

There is excellent evidence that some memories are reactivated during sleep and that this might return them to a labile state^{9,10}.

The crucial question, however, is whether reconsolidation occurs in humans. Memory reconsolidation in people exposed to electroconvulsive-shock therapy was reported some 30 years ago¹¹, but it was not clear whether the phenomenon could be detected under more natural conditions. Walker *et al.*³ have now demonstrated exactly this. They looked at whether consolidated motor memories, when reactivated, would return to a labile state that would again be sensitive to interference. Subjects trained on the first finger-tapping exercise on day 1 performed with greater speed and accuracy when they were retested on days 2 and 3. But if, on day 2, they were given the second exercise immediately after the retest of the first, all improvements in the accuracy of the first exercise were reversed by day 3. Strikingly, this interference was prevented if the first exercise was not retested on day 2 — in other words, if this memory was not reactivated.

These results show that reactivating a consolidated motor memory returns it to a labile state that is once again sensitive to interference. Walker and colleagues show that new memories and reactivated memories respond in qualitatively similar ways when challenged using interference, so the most parsimonious interpretation of the data is that the reactivation of the consolidated motor memory caused it to undergo reconsolidation. The interference-induced block to reconsolidation might represent the first convincing demonstration of the erasure of a consolidated memory in humans.

The finding that reconsolidation occurs in humans is a landmark discovery. But Walker *et al.* also demonstrate that not all memories are equal. Although the accuracy of the finger-tapping tests was affected by interference, the speed with which they were performed was not. So perhaps not all memories undergo reconsolidation. Alternatively, the reconsolidation of different memories might be affected by interference in different ways. Understanding the intricacies of this phenomenon is crucial if we are to exploit its clinical potential to address psychological disorders such as post-traumatic stress or addiction. ■

Karim Nader is in the Department of Psychology, McGill University, Montreal, Quebec H3A 1B1, Canada.

e-mail: nader@hebb.psych.mcgill.ca

1. Nader, K. *Trends Neurosci.* **26**, 65–72 (2003).
2. Fenn, K. M., Nusbaum, H. C. & Margoliash, D. *Nature* **425**, 614–616 (2003).
3. Walker, M. P., Brakefield, T., Hobson, J. A. & Stickgold, R. *Nature* **425**, 616–620 (2003).
4. Müller, G. E. & Pilzecker, A. Z. *Psychol. Suppl.* **1**, 1–288 (1900).
5. Kandel, E. R. *Science* **294**, 1030–1038 (2001).
6. Siegel, J. M. *Science* **294**, 1058–1063 (2001).
7. Karni, A. *et al. Proc. Natl Acad. Sci. USA* **95**, 861–868 (1998).
8. Mednick, S. C. *et al. Nature Neurosci.* **5**, 677–681 (2002).
9. Wilson, M. A. & McNaughton, B. L. *Science* **265**, 676–679 (1994).
10. Spear, N. E. & Gordon, W. C. in *Sleep, Dreams and Memory* (ed. Fishbein, W.) 183–203 (Spectrum, New York, 1981).
11. Rubin, R. D., Fried, R. & Franks, C. M. in *Advances in Behavior Therapy, 1968* (eds Rubin, R. D. & Franks, C. M.) 37–44 (Academic, New York, 1969).



100 YEARS AGO

Certain letters have appeared in *NATURE* upon the bearing of the properties of radium upon the cosmical time scale. These letters are based on the assumption that radium, or some equally active body, exists in the sun and contributes materially to the output of solar energy. If this assumption were true, we ought, I think, to be able to detect the rays peculiar to radio-active bodies on the surface of the earth — they should bear some proportion to the great stream of light and heat waves which reaches us. Now a solution of iodoform in chloroform is very sensitive to the β and γ rays. A purple coloration is produced by the rays from 5mg. of radium bromide even after filtering through 1cm. of lead. On the other hand, I find that direct sunlight (if heating be obviated) has no action when the thinnest screen is interposed even after many days. Some of my solutions are now nearly two months old, and they have been exposed in light-tight cardboard boxes to such sunshine as has reached us during that period. They are quite unchanged. It is, of course, possible that the stream of rays needs to be above a certain critical density in order to decompose the iodoform, but in any case, my experiments prove that the β and γ rays reach us at most in only faint quantities from the sun.

From *Nature* 8 October 1903.

50 YEARS AGO

During a short survey of Milford Haven (July 27–30, 1953) for the purpose of assessing the possibility of reviving the oyster fishery, six American slipper limpets (*Crepidula fornicata*) were found in or near Pennar Gut, the estuary of the Pembroke River. *Crepidula*, a serious pest of oysters, had previously been recorded only from the east and south coasts of Britain... Pennar Gut was used in the years following the Second World War for laying up both merchant and naval ships, and is adjacent to Pembroke Dock. There is a ship-breaking yard on the opposite shore near Milford Haven. Evidence has been adduced to show that a major factor in the spread of *Crepidula* has been the movement of laid-up naval and merchant ships, which may remain for several years in infested areas on the east and south coasts of Britain and later be transferred for refitting or breaking-up to other areas.

From *Nature* 10 October 1953.

Quantum physics

How to make a Schrödinger's cat

Phys. Rev. Lett. **91**, 130401 (2003)

Can a macroscopic object be placed in a quantum superposition — like Schrödinger's notorious cat, both dead and alive? William Marshall *et al.* have devised a scheme that comes close. Their 'cat' is not alive, nor exactly macroscopic; but it is far bigger than the atomic-scale entities with which superpositions are generally associated.

The researchers claim that a microscopic mirror, 10 μm square and weighing around 5×10^{-12} kg, could be placed in a superposition such that it can be considered to occupy two positions simultaneously. The mirror is attached to a cantilever arm of the sort used in scanning probe microscopes, and forms one half of an optical cavity in which photons bounce back and forth. By coupling this cavity to an interferometer in which a light beam is split along two paths, a single photon placed in a superposition in the interferometer creates a superposition of the mirror. This would be revealed by the disappearance and reappearance of interference as the mirror cycles between its initial state and a quantum superposition.

Marshall *et al.* estimate that, for a mirror containing around 10^{14} atoms, the apparatus can be made sufficiently sensitive with existing technology. It would have to be cooled to around 2 mK, and the mirror itself further cooled using laser cooling to initialize it in the ground state. The technical demands are challenging, but seem feasible.

Philip Ball

Neuroscience

Transport delays

Neuron **40**, 41–52 (2003)

Huntington's disease is an incurable, degenerative brain disorder. Sufferers carry an abnormal huntingtin protein, which contains longer repetitive stretches of the amino acid glutamine than normal. The protein is found inside nerve cells, often suspiciously close to vesicles — miniature suitcases that transport molecules along nerve-cell axons.

Györgyi Szebenyi *et al.* hypothesized that the mutant protein might disrupt this transport system. The team treated axons from giant squids with glutamine-rich fragments of the huntingtin protein. Vesicular transport slowed by more than 50%. The same treatment also stopped intact human nerve cells from extending axons normally.

In the brain, these failures may prevent

essential molecules from reaching their destination. As a result, neurons may malfunction and die, the authors speculate. Similar mechanisms may operate in other polyglutamine diseases, such as Kennedy's disease or spinobulbar muscular atrophy. Drugs that facilitate axonal transport may help to slow or block neurodegenerative diseases of this type.

Helen R. Pilcher

Mineralogy

Oresome microbes

Geology **31**, 913–916 (2003)

Bacteria created massive zinc-ore deposits in Nevada, according to Thomas M. Bawden *et al.* They have found what seems to be the first instance of 'supergene' formation of sphalerite (zinc sulphide, the principal commercial form of zinc ore). Supergene deposits are formed by the chemical or biological enrichment of an ore in a near-surface environment. Previously, it was thought that supergene enrichment of sphalerite was thermodynamically unfavourable and unlikely to supply deposits suitable for mining.

Bacteria have a well-known capacity to induce mineral formation, but it is rare for this to happen on a scale big enough to produce economically viable mineral deposits. In the gold-rich strata studied by Bawden and colleagues — the Mike deposit in northern Nevada, which dates to the late Eocene (about 35 million years ago) — sulphate-reducing bacteria apparently generated more than 400,000 tonnes of zinc ore during the subsequent Miocene epoch (24–25 million years ago). The involvement of bacteria is revealed by depletion of ^{34}S in the sphalerite, a characteristic signature of a biogenic origin.

Philip Ball

Planetary science

Less of an impact

Geology **31**, 869–872 (2003)

Meteorite impacts do not cause volcanic eruptions, according to B. A. Ivanov and H. J. Melosh. Their views run contrary to a theory popular among geologists, but Ivanov and Melosh say the theory doesn't hold water (or lava), and they have the calculations to support their case.

Since the first glimpse of lava-filled craters on the near side of the Moon in the 1960s, scientists have speculated that volcanism can be caused by decompression melting of crustal rock deep beneath impact craters. The lunar lava proved to be much younger than the impacts, but the idea persisted.

Using numerical simulations, Ivanov and Melosh calculated that a crater would need to be at least 500 km in diameter, twice the size of any known crater on Earth, before any molten material would reach the surface, even in relatively hot oceanic crust. Earth is thought to experience an impact of this size only once or twice every billion years. And an impact on land would have to be at least 1,200 km across. Even colossal impacts on other planets, such as the one that created the 2,000-km Hellas Basin on Mars (pictured), failed to trigger volcanism.

Betsy Mason

Genetics

Dyslexic inheritance

Proc. Natl Acad. Sci. USA **100**, 11553–11558 (2003)

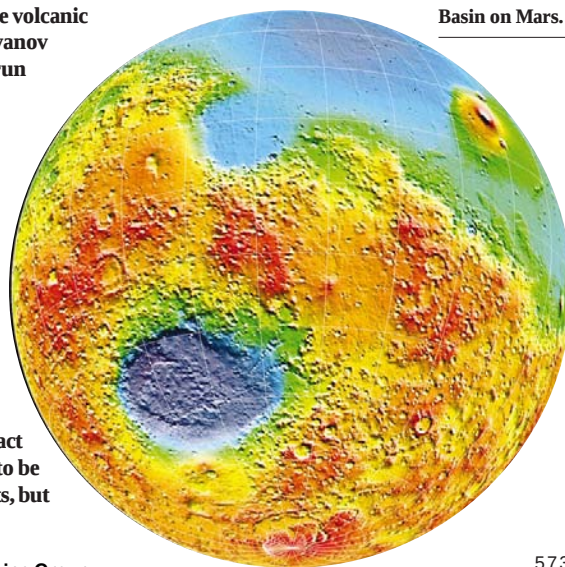
Mikko Taipale and co-workers have pin-pointed a gene that, when disrupted, is implicated in dyslexia. It is normally found on the long arm of chromosome 15, and is called *DYX1C1*. The protein it encodes is predicted to be 420 amino acids long; its function is unknown.

The starting point for Taipale *et al.* was a family in which the father and two daughters all suffered from dyslexia. It turned out that all three carried a translocation of *DYX1C1* between chromosome 15 and chromosome 2. Further investigation of the gene in other groups of people with and without dyslexia revealed two changes in gene sequence that were associated with the disorder to greater or lesser extents.

The next steps will be to find out whether disrupted *DYX1C1* occurs in broader samples of dyslexic patients, and whether it is involved in a more general learning disorder. Overall, the genetics of predisposition to dyslexia is likely to be fiendishly complex — many variants of many genes might be involved, and some may be specific to different languages.

Tim Lincoln

Hell of a hit — the Hellas Basin on Mars.



Gas-bubble lesions in stranded cetaceans

Was sonar responsible for a spate of whale deaths after an Atlantic military exercise?

There are spatial and temporal links between some mass strandings of cetaceans — predominantly beaked whales — and the deployment of military sonar^{1–3}. Here we present evidence of acute and chronic tissue damage in stranded cetaceans that results from the formation *in vivo* of gas bubbles, challenging the view that these mammals do not suffer decompression sickness. The incidence of such cases during a naval sonar exercise indicates that acoustic factors could be important in the aetiology of bubble-related disease and may call for further environmental regulation of such activity.

Fourteen beaked whales were stranded in the Canary Islands close to the site of an international naval exercise (Neo Tapon, 2002) held on 24 September 2002. Strandings began about 4 hours after the onset of mid-frequency sonar activity. We necropsied eight Cuvier's beaked whales (*Ziphius cavirostris*), a Blainville's beaked whale (*Mesoplodon densirostris*) and a Gervais' beaked whale (*Mesoplodon europaeus*), six of which were very fresh. These animals showed severe, diffuse vascular congestion and marked, disseminated microvascular haemorrhages associated with widespread fat emboli within vital organs. Intravascular bubbles were present in several organs, although definitive evidence of gas embolism *in vivo* is difficult to determine after death⁴. No pathogenic bacteria were isolated from the carcasses.

These lesions are consistent with acute trauma due to *in vivo* bubble formation resulting from rapid decompression (as occurs in decompression sickness)^{4,5}. Bubble formation in response to sonar exposure might result from behavioural changes to normal dive profiles (such as accelerated ascent rate), causing excessive nitrogen supersaturation in the tissues (as occurs in decompression sickness); alternatively, bubble formation might result from a physical effect of sonar on *in vivo* bubble precursors (gas nuclei) in nitrogen-supersaturated tissues^{6,7}.

The beaked whales found in the Canary Islands are not the only stranded cetaceans to provide evidence of bubble-associated tissue injury. In strandings that occurred in Britain between October 1992 and January 2003, three out of 24 Risso's dolphins (*Grampus griseus*), three out of 342 common dolphins (*Delphinus delphis*) and one out of 1,035 harbour porpoises (*Phocoena phocoena*) necropsied, as well as the only Blainville's beaked whale studied, contained gas bubbles in their blood vessels and gas-filled cavities in their parenchymous organs.

The livers of these animals were the most consistently affected organ, with macroscopic gas-filled cavities (diameter, 0.2–6.0 cm) occupying 5–90% of the volume (Fig. 1) and having variable degrees of fibrotic encapsulation. Intrahepatic spherical non-staining cavities (gas bubbles) of diameter 50–750 µm were associated with compression of hepatic tissue, distension of portal blood vessels, and sometimes with haemorrhage, acute hepatocellular necrosis or fibrosis, indicating that this damage was inflicted before death. One of the *D. delphis* specimens also had bilateral acute renal infarcts associated with gas bubbles.

The cavity lesions described here are new to marine-mammal pathology. Their presence in fresh carcasses in the absence of bacterial isolates, and the apparent progression through the stages of pericavitary fibrosis, are inconsistent with decompositional bubbles from bacterial activity. The coexistence of ante mortem gas bubbles and gas-filled fibrosed cavities suggests that *in vivo* bubble formation and expansion is the proximate aetiology of this disease process. Bubble formation may either have initiated in the liver and kidneys ('autochthonous' bubble formation) or in other tissues (fatty tissue, for example) before haematogenous transfer to the liver and kidneys as gas emboli.

Nitrogen bubbles and emboli can develop in decompression sickness in humans and

experimental animals as a result of expansion of pre-existing gas nuclei within nitrogen-supersaturated tissues⁵. Anatomical, physiological and behavioural adaptations may mitigate against *in vivo* formation of nitrogen bubbles in marine mammals^{8–12}, although there is empirical evidence of nitrogen supersaturation in cetaceans⁸. Some deep-diving species are predicted to undergo up to 300% nitrogen tissue supersaturation⁷. Static diffusion in nitrogen-supersaturated tissues is therefore a plausible mechanism for bubble development and is consistent with a greater prevalence of cases in deep-diving species such as Risso's dolphins and beaked whales.

Further investigation is needed into the physical and behavioural effects on cetaceans exposed to sonar, and the relation of these effects to bubble growth *in vivo* and to strandings. Necropsies should aim to compare fat and gas emboli in stranded cetaceans suspected of having been exposed to sonar with results from unexposed stranded controls. In a wider conservation sense, our findings need to be taken into account in considering the regulation and limitation of the adverse impact of anthropogenic sonar on cetaceans.

P. D. Jepson*, **M. Arbelo†**, **R. Deaville***, **I. A. P. Patterson‡**, **P. Castro†**, **J. R. Baker§**, **E. Degollada†**, **H. M. Ross†**, **P. Herráez†**, **A. M. Pocknell***, **F. Rodríguez†**, **F. E. Howie||**, **A. Espinosa†**, **R. J. Reid†**, **J. R. Jabert†**, **V. Martín†**, **A. A. Cunningham***, **A. Fernández†**

*Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

†Histology and Pathology Unit, Institute for Animal Health, Veterinary School, Montana Cardones-Arucas, University of Las Palmas de Gran Canaria, Gran Canaria, Spain
e-mail: afernandez@dmor.ulpgc.es

‡Wildlife Unit, SAC Veterinary Science Division (Inverness), Drummondhill, Stratherrick Road, Inverness, IV2 4JZ, UK

§Department of Veterinary Pathology, University of Liverpool, Leahurst, Neston, Wirral CH64 7TE, UK
||SAC Veterinary Science Division (Edinburgh), Allan Watt Building, Bush Estate, Penicuik, Edinburgh EH26 0QE, UK

1. Simmonds, M. P. & Lopez-Jurado, L. F. *Nature* **337**, 448 (1991).
2. Frantz, A. *Nature* **392**, 29 (1998).

3. US Dept Commerce & US Navy Joint Interim Report: Bahamas Marine Mammal Stranding Event of 15–16 March 2000
http://www.nmfs.noaa.gov/prot_res/overview/Interim_Bahamas_Report.pdf (2001).

4. Knight, B. *Forensic Pathology* 2nd edn (Arnold, London, 1996).

5. Francis, T. J. R. & Mitchell, S. J. in *Bennett and Elliott's Physiology and Medicine of Diving* (eds Brubakk, A. O. & Neuman, T. S.) 530–556 (Saunders, Philadelphia, 2003).

6. Crum, L. & Mao, Y. J. *Acoust. Soc. Am.* **99**, 2898–2907 (1996).

7. Houser, D. S., Howard, R. & Ridgway, S. H. *J. Theor. Biol.* **213**, 183–195 (2001).

8. Ridgway, S. H. & Howard, R. *Science* **206**, 1182–1183 (1979).

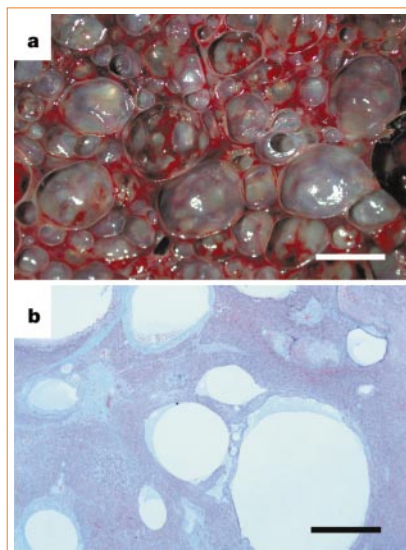


Figure 1 Gas-filled cavities in the liver of a stranded common dolphin (*Delphinus delphis*). **a**, Cut surface of the liver, showing that cavity lesions have extensively replaced the normal tissue. Scale bar, 10 mm. **b**, Photomicrograph of liver section, showing multiple cavities (gas bubbles) within the portal tracts and hepatic parenchyma. Scale bar, 750 µm.

9. Pongonis, P. J., Kooyman, G. L. & Ridgway, S. H. in *Bennett and Elliott's Physiology and Medicine of Diving* (eds Brubakk, A. O. & Neuman, T. S.) 211–226 (Saunders, Philadelphia, 2003).
10. Harrison, R. J. & Tomlinson, D. W. *Proc. Zool. Soc. Lond.* **126**, 205–233 (1956).
11. Ridgway, S. H. & Howard, R. *Science* **216**, 651 (1982).
12. Falke, K. J. *et al. Science* **229**, 556–558 (1985).

Competing financial interests: declared none.

Immunology

Hepatitis A virus link to atopic disease

Atopic diseases, including asthma, allergic rhinitis and atopic dermatitis, are caused by both environmental and genetic factors. Here we show that infection by hepatitis A virus (HAV) may protect individuals from atopy if they carry a particular variant of the gene that encodes TIM-1 (also known as HAVcr-1) — the cell-surface receptor used by HAV to infect human cells¹. Exposure to HAV is associated with poor hygiene, large family size and attendance at day-care centres, all factors that are also inversely associated with atopy^{2–6}. Our discovery indicates that interaction between HAV and TIM-1 genotype may contribute to the aetiology of atopic diseases, and provides a mechanism to account for the hygiene hypothesis.

Using a congenic positional cloning strategy, we identified *TIM-1* as a candidate gene for atopy and asthma in a region of mouse chromosome 11, which is homologous to a segment of human chromosome 5q31–33 that has been linked to atopy^{7,8}. *TIM-1* is expressed by activated CD4⁺ T cells during the development of helper-T-cell (Th2) responses and regulates cytokine production⁹. We therefore investigated whether the interaction between HAV and TIM-1 on lymphocytes can modify T cells in a way that protects against atopy, and whether polymorphisms in *TIM-1* can alter susceptibility to atopy⁷.

By sequencing complementary DNA from human lymphocytes, we identified a six-amino-acid insertion (ins) at residue 157, termed 157insMTTTPV (one-letter amino-acid notation), as well as two single-amino-acid changes, 195delT (where 'del' signifies a deletion) and A206T. The insertion 157insMTTTPV is located at the centre of an extracellular mucin-like region that is required for efficient HAV uncoating⁹ and, because 157insMTTTPV lengthens this critical region by 12–14%, this variation may affect the efficiency of viral entry (Fig. 1).

To determine the effect of the insertion 157insMTTTPV on the occurrence of atopy, we carried out a cross-sectional study of 375 individuals who were evaluated by history and tested serologically for atopy and prior HAV infection. To correct for potentially confounding effects of population admixture, we used stratified Mantel–Haenszel χ^2

tests to quantify the association between atopy and 157insMTTTPV in the total sample. We found that HAV seropositivity protects against atopy, but only in individuals with the 157insMTTTPV variant of TIM-1 ($P = 0.0005$; Table 1).

The protective effects of HAV therefore depend upon a common *TIM-1* allele that is carried by 63% of Caucasians, 46% of Asians and 64% of African Americans in this population (see supplementary information). As allelic variation in *TIM-1* does not affect HAV-infection rates in our population ($\chi^2 = 1.567$, $P = 0.211$), we conclude that the interaction of HAV with *TIM-1* genotype seen here is not due to variation in the rate of seroconversion following HAV exposure.

Before 1970, the seroprevalence of antibodies against HAV approached 100% in Western countries⁴, and infection with HAV may have protected many individuals against atopy³. However, modernization has led to a reduction in average family size and significant improvements in public health, causing anti-HAV seroprevalence to fall to 25–30%, while the prevalence of atopic disease has doubled⁴. Our finding that *TIM-1* is associated with atopy in HAV-seropositive individuals indicates that exposure to a specific pathogen may influence the expression of atopy — so a declining prevalence of HAV infection could contribute to an increase in atopy by association with *TIM-1*. It will be necessary to determine whether HAV exposure must occur during childhood to have a protective effect, whether HAV can mitigate the severity of existing atopic disease, and whether HAV vaccination can reproduce the effects of natural HAV infection.

Jennifer J. McIntire*, **Sarah E. Umetsu***, **Claudia Macaubas***, **Elizabeth G. Hoyte***, **Cengiz Cinnioglu†**, **Luigi L. Cavalli-Sforza†**, **Gregory S. Barsh†**, **Joachim F. Hallmayer‡**, **Peter A. Underhill†**, **Neil J. Risch†**, **Gordon J. Freeman§**, **Rosemarie H. DeKruyff***, **Dale T. Umetsu***

*Division of Immunology and Allergy, Department of Pediatrics, †Department of Genetics, and ‡Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305-5208, USA

Table 1 TIM-1 insert and protection against atopy

HAV status	157insMTTTPV genotype	Total	Atopic	Non-atopic	χ^2 (P)	Odds ratio (95% CI)
Seronegative (n = 198)	Insertion	120	83 (69%)	37 (31%)	0.463 (0.496)	1.285 (0.708–2.439)
	No insertion	78	50 (64%)	28 (36%)		
Seropositive (n = 123)	Insertion	65	31 (48%)	34 (52%)	11.978 (0.0005)	0.257 (0.116–0.570)
	No insertion	58	46 (79%)	12 (21%)		

Comparison of allele distributions across subjects using the Cochran–Mantel–Haenszel χ^2 test with racial stratification, two-sided tests of significance (P), and number of subjects within each genotype (see supplementary information). The six-amino-acid insertion 157insMTTTPV is associated with atopy in individuals seronegative for hepatitis A virus but not in seropositive individuals. HAV does not independently affect atopy ($\chi^2 = 0.513$, $P = 0.474$). Subgroup analysis of Caucasians and Asians confirms this association in both groups ($P = 0.024$ and $P = 0.036$, respectively), and Breslow–Day tests of the homogeneity of the odds ratios demonstrate no significant difference between racial strata (see supplementary information). This table excludes data from 54 individuals with an intermediate atopic phenotype (see supplementary information).

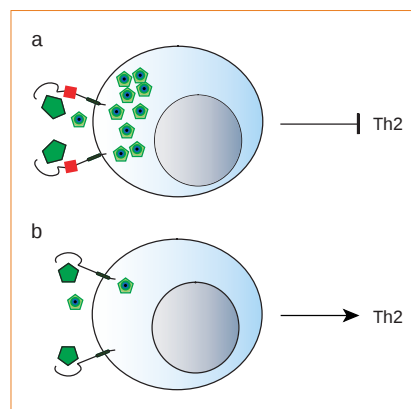


Figure 1 Possible mechanisms of interaction between the cell-surface TIM-1 receptor and hepatitis A virus (HAV), and the effect of HAV on cytokine production. **a**, A variant of TIM-1 ('hook') that carries a 6-amino-acid insert (red), 157insMTTTPV, may increase binding of HAV (large pentagons) to the receptor, thereby enhancing HAV viral uncoating (small pentagons) and infection of TIM-1-expressing T cells. This could lead to deletion of certain lymphocyte subsets, such as Th2 cells, or reduce Th2-cell differentiation, causing a reduction in atopy and asthma. **b**, Alternatively, HAV may bind less efficiently to the form of TIM-1 without the insertion, resulting in less HAV infection and hence more Th2-cell development and more atopy. The mechanism that underpins this interaction between TIM-1, its 157insMTTTPV region and HAV could relate to viral uncoating⁹, the extent and duration of HAV viraemia¹⁰, or to a direct effect on Th1/Th2-subset differentiation.

e-mail: umetsu@stanford.edu

§Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts 02115, USA

1. Feigelsstock, D., Thompson, P., Mattoo, P., Zhang, Y. & Kaplan, G. G. *J. Virol.* **72**, 6621–6628 (1998).
2. Matricardi, P. M. *et al. Br. Med. J.* **314**, 999–1003 (1997).
3. Matricardi, P. M., Rosmini, F., Panetta, V., Ferrigno, L. & Bonini, S. *J. Allerg. Clin. Immunol.* **110**, 381–387 (2002).
4. Bach, J. F. N. *Engl. J. Med.* **347**, 911–920 (2002).
5. Strachan, D. P. *Br. Med. J.* **299**, 1259–1260 (1989).
6. Kramer, U., Heinrich, J., Wjst, M. & Wichmann, H. E. *Lancet* **353**, 450–454 (1999).
7. McIntire, J. J. *et al. Nature Immunol.* **2**, 1109–1116 (2001).
8. Marsh, D. G. *et al. Science* **264**, 1152–1156 (1994).
9. Silberman, E. *et al. J. Virol.* **77**, 8765–8774 (2003).
10. Bower, W. A., Nainan, O. V., Han, X. & Margolis, H. S. *J. Infect. Dis.* **182**, 12–17 (2000).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Smad-dependent and Smad-independent pathways in TGF- β family signalling

Rik Derynck¹ & Ying E. Zhang²

¹Departments of Growth and Development, and Anatomy, Programs in Cell Biology and Developmental Biology, University of California at San Francisco, San Francisco, California 94143-0640, USA

²Laboratory of Cellular and Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-4254, USA

Transforming growth factor- β (TGF- β) proteins regulate cell function, and have key roles in development and carcinogenesis. The intracellular effectors of TGF- β signalling, the Smad proteins, are activated by receptors and translocate into the nucleus, where they regulate transcription. Although this pathway is inherently simple, combinatorial interactions in the heteromeric receptor and Smad complexes, receptor-interacting and Smad-interacting proteins, and cooperation with sequence-specific transcription factors allow substantial versatility and diversification of TGF- β family responses. Other signalling pathways further regulate Smad activation and function. In addition, TGF- β receptors activate Smad-independent pathways that not only regulate Smad signalling, but also allow Smad-independent TGF- β responses.

The TGF- β superfamily comprises TGF- β s, bone morphogenetic proteins (BMPs), activins and related proteins. These proteins were identified mainly through their roles in development; they regulate the establishment of the body plan and tissue differentiation through their effects on cell proliferation, differentiation and migration. The growth-inhibitory effect of TGF- β signalling in epithelial cells explains its role as a tumour suppressor in carcinomas, although TGF- β expression by tumour cells contributes to cancer progression as well. The current model of induction of signalling responses by TGF- β -related factors (Fig. 1) is a linear signalling pathway from the type II to the type I receptor kinase to Smad activation, resulting in ligand-induced transcription^{1–3}.

Although there are considerably fewer receptors and Smads than there are ligands, a greater versatility of signalling is possible than might be expected. Combinatorial interactions of type I and type II receptors and Smads in oligomeric complexes allow substantial diversity, and are complemented by the many sequence-specific transcription factors with which Smads cooperate, resulting in context-dependent transcriptional regulation. Other signalling pathways help to define the responses to TGF- β factors, and it is increasingly apparent that TGF- β -related proteins activate not only Smads but also other signalling pathways. These pathways regulate Smad-mediated responses, yet also induce Smad-independent responses. Here we summarize recent progress toward understanding the signalling mechanisms of TGF- β -related factors through Smad-dependent and Smad-independent pathways.

Versatility in receptor interactions and ligand binding

The functional complex of TGF- β family receptors at the cell surface consists of two 'type II' and two 'type I' transmembrane serine/threonine kinase receptors^{1–4} (Fig. 2). The latter have a characteristic Gly-Ser (GS) sequence upstream from the kinase domains. In mammals, only five type II receptors and seven type I receptors have been identified^{2–4}, whereas 29 ligands have been found. In the absence of ligand, type II and type I receptors exist as homodimers at the cell surface. TGF- β 1, TGF- β 3 and activins bind their type II receptors without needing a type I receptor, whereas BMP-2, BMP-4 and BMP-7 bind primarily to their type I receptors, BMP-RIA or BMP-RIB, although heteromeric BMP receptor complexes provide higher-affinity ligand binding⁴. Some TGF- β ligands—for example, TGF- β 2—interact only with type II–type I receptor combinations, suggesting that heteromeric receptor complexes form in the absence

of ligand, consistent with the inherent affinity of the receptors for each other^{4,5}. Thus, the low-affinity heteromeric receptor complex may provide a surface for ligand binding that conformationally stabilizes the complex. Ligand binding to either homomeric receptor dimer is not sufficient to activate signalling. By contrast, activation of the type I receptor kinase, and consequent signalling, requires phosphorylation of its GS domain by the type II receptor in the heteromeric complex^{1,4}. Although ligand binding to type II receptor dimers may induce cytoplasmic-domain autophosphorylation, type II receptor signalling in the absence of type I receptors has not been reported.

Combinatorial interactions in the tetrameric receptor complex allow differential ligand binding or differential signalling in response to the same ligand⁴ (Fig. 2). One receptor combination often binds different ligands, and patterns of ligand and receptor expression often dictate which receptors are activated. For example, the type II receptors ActRII and ActRIIB can combine with the type I receptor ActRI/ALK4 and mediate activin signalling, whereas their interactions with BMP-RIA or BMP-RIB allow BMP binding and signalling instead. The BMP type II receptor BMP-RII can combine with three type I receptors, BMP-RIA, BMP-RIB and ActRI/ALK2, to bind several BMPs and mediate BMP signalling. Thus, a ligand can induce different signalling pathways depending on the composition of the receptor complex. For example, T β RII interacts not only with the 'classical' type I receptor T β RI/ALK5, which activates Smad2 and Smad3, but also with ALK1, which activates Smad1 and Smad5. Differential activation of either receptor in endothelial cells induces different responses to TGF- β , suggesting that a balance in their activation controls the state of the endothelium⁶. In addition, ActRI/ALK2 activation by TGF- β has been implicated in TGF- β -induced epithelial-to-mesenchymal differentiation^{7,8}.

Further complexity is imposed by heterodimeric TGF- β ligands—for example, inhibins and heteromeric BMPs—which may need to bind to heterodimeric type II or type I receptor combinations. Also, accessory proteins enhance or modify ligand-binding specificity. For example, Nodal acts through ActRIIB and ActRI/ALK4 (the activin receptor complex), activating Smad2, but efficient ligand binding and signalling require association of fucosylated Cripto, a TGF- α -like transmembrane protein, with ALK4 (ref. 9). Also, betaglycan and endoglin provide high-affinity TGF- β presentation to the signalling TGF- β receptor complex. Thus, cell-surface expression of betaglycan may regulate the TGF- β response¹⁰, and the vascular abnormalities resulting from impaired endoglin func-

tion are probably explained by defective signalling in endothelial cells by TGF- β ligands that interact with endoglin¹¹.

Smads are structurally related signalling effectors

There are eight vertebrate Smads: Smad1 to Smad8. Smad2 and Smad3 are activated through carboxy-terminal phosphorylation by the TGF- β and activin receptors T β RI and ActRIB, whereas Smad1, Smad5 and Smad8 are activated by ALK-1, ALK-2, BMP-RIA/ALK-3 and BMP-RIB/ALK-6 in response to BMP¹⁻⁴ or other ligands. These receptor-activated Smads (R-Smads) are released from the receptor complex to form a heterotrimeric complex of two R-Smads and a common Smad4, and translocate into the nucleus. *Xenopus* has two Smad4s, Smad4 α and Smad4 β , encoded by two genes¹². Although ubiquitously involved in Smad-mediated transcription, Smad4 is not essential for TGF- β responses: some TGF- β responses occur in the absence of Smad4 and some Smad4-deficient cell lines have a limited responsiveness to TGF- β ¹³. The structurally divergent Smad6 and Smad7 act as 'inhibitory' Smads¹⁻³. R-Smads and Smad4 contain a conserved MH1 and C-terminal MH2 domain, flanking a divergent middle 'linker' segment¹⁻³ (Fig. 3). Inhibitory Smads lack a recognizable MH1 domain, but have an MH2 domain. The MH2 domain has limited structural similarity to the phosphopeptide-binding domain FHA¹⁴. Both the MH1 and the MH2 domains can interact with select sequence-specific transcription factors, whereas the C terminus of the R-Smads interacts with and recruits the related coactivators CREB-binding protein (CBP) or p300 (refs 1-3). With the exception of Smad2, the MH1 domains of

Smads can bind DNA, whereas the MH2 domains mediate Smad oligomerization and Smad-receptor interaction (see Fig. 3).

Regulation of Smad levels

Although differentially controlled during development, R-Smads and Smad4 are expressed in most, if not all, cell types. In contrast to R-Smad expression, expression of the inhibitory Smad6 or Smad7 is highly regulated by extracellular signals. Induction of Smad6 and Smad7 expression by BMP and TGF- β represents an auto-inhibitory feedback mechanism for ligand-induced signalling¹⁻³. Accordingly, the downregulation of Smad6 and Smad7 expression during adipocyte differentiation may result from concomitant loss of autocrine TGF- β and BMP signalling¹⁵. Activation of the epidermal growth factor (EGF) receptor and possibly other tyrosine kinase receptors, interferon- γ signalling through STAT (signal transducer and activator of transcription) proteins, and activation of NF- κ B by tumour-necrosis factor- α , also induce Smad7 expression, leading to inhibition of TGF- β signalling¹⁻³ (Fig. 4).

Ubiquitin-proteasome-mediated degradation controls the levels of Smads post-translationally. The HECT (homologous to the E6-AP carboxy terminus) family E3 ubiquitin ligases, Smurf1 (Smad-ubiquitination-regulatory factor 1) and Smurf2, antagonize TGF- β family signalling by interacting with R-Smads and targeting them for degradation¹⁶. Smurf-mediated degradation thus controls R-Smad levels and the sensitivity of cells to incoming signals. Smurf1 interacts with Smad1 and Smad5, thereby affecting BMP responses¹⁷, whereas Smurf2 interacts more broadly with different R-Smads, allowing interference with BMP and TGF- β /activin signalling^{16,18,19}. Nevertheless, *Xenopus* embryo assays indicate that Smurf1 and Smurf2 primarily target the BMP pathway^{17,18}.

Proteasomal degradation also regulates the R-Smad levels after translocation into the nucleus. Thus, C-terminally phosphorylated Smad2 can undergo ubiquitination, and inhibition of proteasomal degradation enhances its nuclear accumulation²⁰. Nuclear, C-terminally phosphorylated Smad3 can interact with a Ring-finger protein, Roc1, allowing association with an SCF ubiquitin ligase complex consisting of Roc1, Skp1, Cullin1 and β TRCP1/Fbw1a (β -transducin-repeat-containing protein 1) and consequent

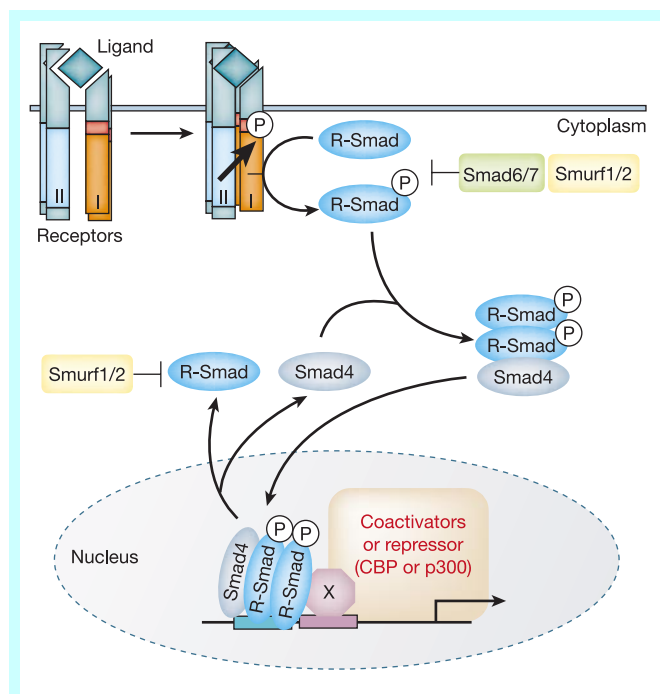
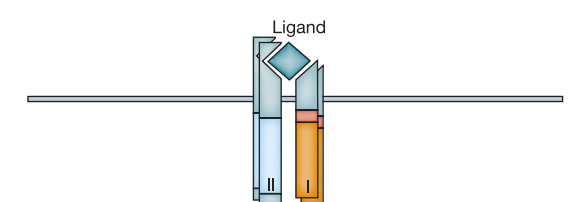


Figure 1 General mechanism of TGF- β receptor and Smad activation. At the cell surface, the ligand binds a complex of transmembrane receptor serine/threonine kinases (types I and II) and induces transphosphorylation of the GS segments (red) in the type I receptor by the type II receptor kinases. The consequently activated type I receptors phosphorylate selected Smads at C-terminal serines, and these receptor-activated Smads (R-Smads) then form a complex with a common Smad4. Activated Smad complexes translocate into the nucleus, where they regulate transcription of target genes, through physical interaction and functional cooperation with DNA-binding transcription factors (X) and CBP or p300 coactivators. Activation of R-Smads by type I receptor kinases is inhibited by Smad6 or Smad7. R-Smads and Smad4 shuttle between nucleus and cytoplasm. The E3 ubiquitin ligases Smurf1 and Smurf2 mediate ubiquitination and consequent degradation of R-Smads, yet can also interact with Smad6/7 and thereby ubiquitinate the type I receptors (not shown).



Type II	Type I	R-Smad
BMPRII	ALK-2 (ActRI) ALK-3 (BMP-RIA) ALK-6 (BMP-RIB)	Smad1, Smad5, Smad8
ActRII, ActRIIB ActRIIB	ALK-4 (ActRIB) ALK-7	Smad2 Smad2
T β RII	ALK-5 (T β RI) ALK-1 ALK-2	Smad2, Smad3 Smad1, Smad5
AMHR	ALK-3 ALK-2 ALK-6	Smad1, Smad5

Figure 2 Combinatorial interactions of type II and type I receptors define the signalling responses. Only the best-documented receptor combinations and their R-Smads are listed.

nuclear export and ubiquitin–proteasome-mediated degradation in the cytoplasm²¹. However, only a small fraction of Smad2 and Smad3, in the absence or presence of TGF- β , is ubiquitinated, and, upon TGF- β signalling, phosphorylated Smad2 or Smad3 can form a stable complex with Smurf2 (ref. 19). Thus, the bulk of nuclear Smad2 or Smad3 is not targeted for degradation, but dephosphorylated and relocated to the cytoplasm^{22,23}.

In contrast to R-Smads, Smad4 is not subjected to ubiquitin-mediated degradation. Instead, sumoylation of Smad4 enhances its stability²⁴. However, some tumour-associated mutations allow ubiquitination and/or decrease the stability of Smad4 (ref. 25). Jab1, the CSN5 subunit of the COP9 signalosome that was initially identified as a coactivator for c-Jun, has been implicated in targeting Smad4 for ubiquitination and degradation²⁶.

Regulation of Smad activation

Ligand-induced interaction of R-Smads with activated type I receptors results in direct phosphorylation of the two distal serines of the C-terminal SSXS motif by the type I receptor kinases. This interaction is specified by sequences in both the receptor and the Smad. The nine-amino-acid L45 loop in the type I receptor kinase domain is the main determinant of receptor signalling and Smad-binding specificity, and interacts directly with the L3 loop in the MH2 domain of the R-Smad. Sequences downstream from the L3 loop also contribute to receptor-binding specificity. The type I receptor's GS sequence, once phosphorylated on serines by the type II receptor, provides an interface with the sequence downstream from the L3 loop, and stabilizes Smad docking and contributes to the Smad-interaction specificity^{1,27}.

TGF- β ligands induce receptor internalization in endosomes, which may be required for efficient TGF- β signalling through Smads^{28,29}. The composition of the heteromeric receptor complex and the interaction of co-receptors not only dictate ligand-binding specificity, but may also confer differential intracellular routing, thereby regulating receptor signalling. Accordingly, receptor-associated proteins may have a role in vesicular trafficking, as well as facilitating TGF- β -induced receptor internalization and Smad recruitment to the receptors^{28,29} (Fig. 4). For example, SARA and Dab2, which are enriched in endosomes and clathrin-coated vesicles, bind both receptors and R-Smads, and promote Smad phosphorylation and TGF- β signalling^{30,31}. TGF- β receptors can also

associate with caveolin, a protein present in plasma-membrane invaginations called caveolae³², and the caveolin-positive lipid-raft compartment is required for receptor turnover and regulates receptor availability and R-Smad activation³⁰. Thus, multiple protein interactions are likely to control subcellular receptor localization and cell-surface receptor availability. These parameters may in turn control the duration of Smad phosphorylation and activation, and thus give rise to qualitatively different responses resulting from different signalling thresholds.

Cytoskeletal proteins also play a part in the localization and signalling of Smads. Unphosphorylated Smad2 and Smad3 bind microtubule filaments, and TGF- β treatment induces their dissociation. Disruption of microtubules with nocodazole increases this dissociation, and enhances Smad2 phosphorylation and activity³³. Smads also interact with filamin, a scaffold for intracellular signalling proteins that crosslinks actin. Cells defective in filamin expression have impaired TGF- β signalling and Smad2 phosphorylation, which can be rescued by ectopic filamin expression³⁴. Finally, TGF- β induces phosphorylation of ELF, a β -spectrin, and its association with Smad3 and Smad4. Lack of ELF results in mislocalization of Smad3/4 and loss of TGF- β -dependent transcription³⁵.

Smad6 and Smad7 also regulate activation of R-Smads^{1–3}. Smad6 and Smad7 inhibit TGF- β family signalling through binding of their MH2 domains to the type I receptor, thus preventing recruitment and phosphorylation of effector Smads^{1–3} (Fig. 4). Smad6 also interferes with the heteromerization of BMP-activated Smads with Smad4, preventing the formation of an effector Smad complex. In addition, recruitment of a complex of Smad7 with Smurf1 or Smurf2 to the type I TGF- β receptor results in receptor ubiquitination by the Smurf proteins and targets the receptor for degradation^{36,37}, possibly at caveolin-containing compartments³⁰, leading to inhibition of R-Smad activation.

Heteromeric interactions of activated Smads

After C-terminal phosphorylation, the R-Smad dissociates from the type I receptor, presumably because of conformational changes that also allow transition from the primarily monomeric, unphosphorylated R-Smad to an oligomeric complex^{38,39}. Activated R-Smads form oligomers in which the phosphorylated C terminus contacts the phosphoserine-binding pockets in the L3 loop region

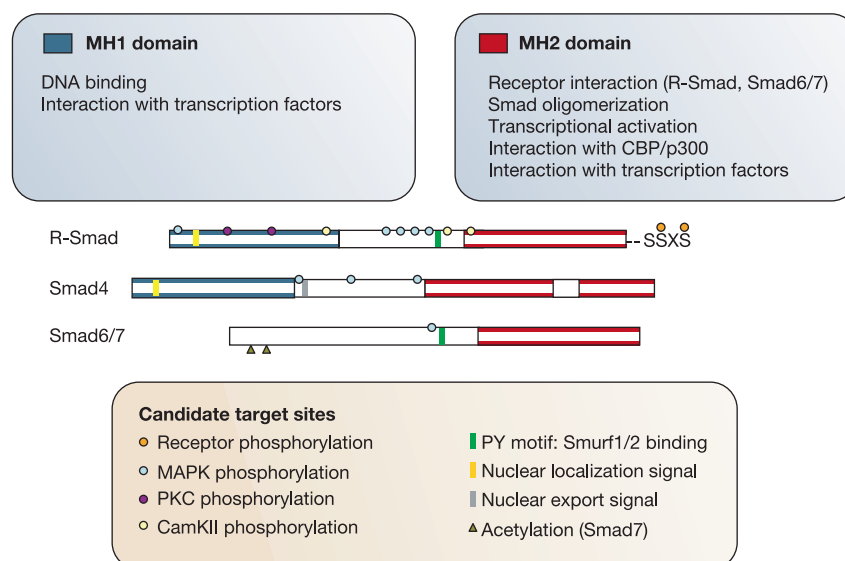


Figure 3 Structural organization and role of the domains of Smads, and candidate target sites for kinase pathways. Such pathways include Erk MAPK and JNK, as well as

CamKII and PKC. The significance of candidate MAPK phosphorylation sites in Smad4 and Smad6/Smad7 is not known.

of a neighbouring R-Smad or Smad4. Endogenous phosphorylated R-Smads may acquire different stoichiometries—heterotrimers with two R-Smads and one Smad4, or heterodimers^{38,39}. Smad heterodimers and heterotrimers may both exist in transcription complexes, depending on the interacting transcription factors; for example, complexes of Smad2–Smad4 with Fast-1 or Fast-3 contain two Smad2s and one Smad4, whereas Smad3–Smad4 complexes at the c-Jun promoter may be heterodimers⁴⁰. Most TGF- β -induced transcription responses are mediated by Smad3 and Smad4, whereas activin responses are mediated by Smad2 and Smad4, suggesting a complex of two Smad3s or two Smad2s, respectively, with Smad4. Some TGF- β responses—for example, induction of p15^{Ink4B} expression—require concomitant activation of Smad2 and Smad3 with Smad4 (ref. 41). The heterotrimeric model allows incorporation of two different R-Smads into the complex and may impose differential signalling thresholds upon gene expression.

Nuclear translocation of Smads

TGF- β receptors remain active for at least 3–4 h after ligand binding, and continuous receptor activation maintains the Smad complexes in the nucleus, where they regulate gene expression²². Nuclear import of a Smad complex follows ‘classical’ nuclear translocation paradigms, established through studies of other proteins. Without ligand stimulation, R-Smads localize in the cytoplasm, whereas Smad4 is distributed in the nucleus and cytoplasm^{22,42,43}. Nuclear import of R-Smads does not require Smad4, although Smad4 cotranslocates with the R-Smads. Nuclear import of Smad1 and Smad3 is conferred by a lysine-rich nuclear localization sequence (NLS) in the MH1 domain that is conserved in all R-Smads⁴³. C-terminal phosphorylation of the MH2 domain and consequent

conformational changes may expose the NLS and allow importin- β binding^{44,45}. In contrast to Smad3, nuclear import of Smad2 may be independent of importin- β , owing to an insertion in its MH1 domain, and may require the MH2 domain instead^{22,23,43,45}. In the nucleus, R-Smads are constantly dephosphorylated, resulting in dissociation of Smad complexes and export of inactive Smads to the cytoplasm^{22,23}. Nucleocytoplasmic shuttling of Smad2 requires its MH2 domain to interact with nucleoporins CAN/Nup214 and Nup153 (ref. 23). In contrast to ligand-dependent import of R-Smads, Smad4 continuously shuttles between the nucleus and cytoplasm owing to the combined activities of a constitutively active NLS in the MH1 domain and a nuclear export signal (NES) in the linker region, whose activity depends on the nuclear transport receptor CRM1 (refs 22, 42). This NES may be masked in the complex with R-Smads, allowing Smad complexes to accumulate in the nucleus. Remarkably, *Xenopus* Smad4 β , which lacks the NES, localizes almost exclusively in the nucleus, whereas Smad4 α , which has an NES, is predominantly cytoplasmic in unstimulated cells^{12,42}.

Smad7 and Smad6 are localized in the nucleus in some cells in the absence of TGF- β ^{46,47}. Their activities in transcription assays^{48,49} and the interaction of Smad7 with CBP/p300 (ref. 50) suggest that they are transcription (co)factors, which might explain their cooperation with TGF- β signalling in inhibiting adipocyte differentiation and stimulating cell proliferation¹⁵. TGF- β induces Smad7 export, whereas BMP induces export of Smad6 (refs 46, 47). TGF- β -induced export of Smad7 may require an interaction with Smurf1 or Smurf2 (ref. 47), whereas BMP-induced cytoplasmic localization of Smad6 may require association with a cytoplasmic retention protein⁴⁶.

Regulation of Smad activity by kinase pathways

C-terminal phosphorylation by the type I receptor is the key event in Smad activation^{1–3}; however, other kinase pathways further regulate Smad signalling (Fig. 3), as suggested by the complex phosphorylation patterns of endogenous Smads. Smad2 phosphorylation and transcription in response to EGF and hepatocyte growth factor, which act through receptor tyrosine kinase receptors, challenge the belief that only TGF- β ligands activate Smads⁵¹.

The Erk mitogen-activated protein kinase (MAPK) pathway, stimulated by the activation of tyrosine kinase receptors and/or Ras, targets R-Smads. Erk MAPK phosphorylates the MH1 domain of Smad2 and the linker segments of Smad1, Smad2 and Smad3 (refs 52, 53). Tyrosine kinase receptor activation and oncogenic Ras inhibit ligand-induced nuclear translocation of activated Smads⁵², which could explain the impaired TGF- β response in some cells with hyperactive Ras signalling. Other studies have not found impaired nuclear translocation of Smads in Ras-transformed cells or in cells with activated MAPK signalling^{51,53,54}. Furthermore, the cooperation between Ras/MAPK and TGF- β signalling in tumour cell behaviour does not seem to be compatible with the defective Smad signalling in Ras-transformed cells⁵⁵. Other kinases may phosphorylate Erk MAPK sites, as suggested by the phosphorylation of these sites in Smad2, but not Smad1, during development⁵⁶.

Phosphorylation of Smads can also result from the activation of MAPK/Erk kinase kinase 1 (MEKK1), which acts downstream from Ras and upstream from growth-factor-induced Erk MAPK and stress-activated SAPK/JNK (c-Jun N-terminal kinase) pathways. MEKK1 activation enhances Smad2 phosphorylation, heteromerization with Smad4, nuclear translocation and transcriptional activity⁵⁷. JNK phosphorylates Smad3 outside its C-terminal SSXS motif, and enhances TGF- β -induced nuclear translocation and transcription⁵⁴. The induction of Erk MAPK and JNK signalling by TGF- β itself (see below) may regulate Smad activation and signalling.

Activation of Ca²⁺/calmodulin-dependent protein kinase II (CamKII) also results in Smad2, Smad3 and Smad4 phosphorylation, inhibits TGF- β -induced nuclear import and transcriptional

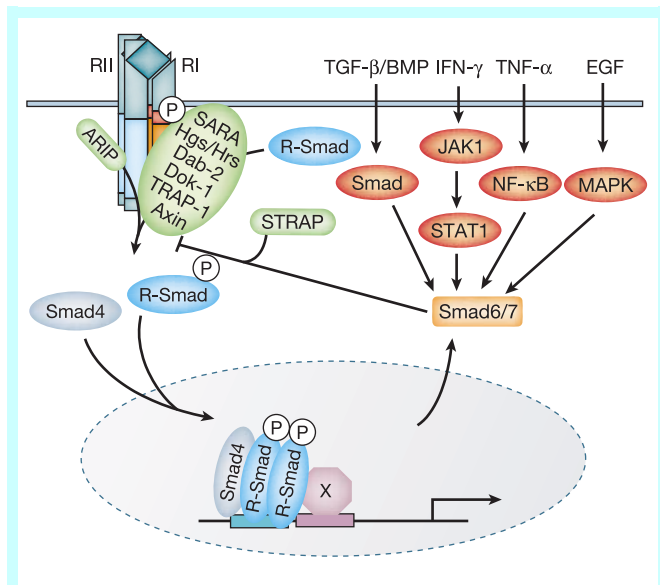


Figure 4 R-Smad activation is regulated by receptor-interacting proteins and Smad6/Smad7. SARA, Hgs/Hrs, Dab2, Dok-1, TRAP-1 (TGF- β -receptor-associated protein), Axin and ARIP (activin-receptor-interacting protein) (green) interact with type I or type II receptors and R-Smads (see refs 1–3 and references therein). SARA or Hrs and Dab2 stabilize the Smad2/Smad3 interactions with TGF- β type I receptors and function in internalization with the endocytic machinery in endosomes. Other proteins, such as the RasGAP-binding protein Dok-1 (ref. 96), the PDZ-domain protein ARIP1 and axin also probably control subcellular localization of receptors and link Smad2/Smad3 to the receptors. TRAP-1, a homologue of the yeast sorting protein Vam6p, interacts with TGF- β or activin type I receptors first, and then with Smad4 upon receptor activation, possibly facilitating Smad4 interaction with activated Smad2 or Smad3. Smad6 and/or Smad7 expression can be induced by several signalling pathways, including TGF- β /BMP signalling through Smads, and attenuates R-Smad activation. STRAP interacts with type I and type II receptors and with Smad7, thus stabilizing the interaction of Smad7 with the receptor complex.

activity of Smad2, and affects Smad heteromerization. CamKII phosphorylates Smad2 in the linker segment and the MH1 domain⁵⁸. The phosphorylation of the MH1 domains of Smad2 and Smad3 by protein kinase C (PKC), which abrogates DNA binding of Smad3, suggests a regulatory role for PKC in Smad-mediated transcription⁵⁹. In contrast to R-Smads, mammalian Smad4 is not regulated by phosphorylation. In *Xenopus*, Smad4 β , but not Smad4 α , is phosphorylated, but the role of this phosphorylation is unknown¹². Smad6 and Smad7 are also phosphorylated independently of TGF- β stimulation. Mutation of a serine that is phosphorylated in Smad7 alters its transcriptional activity when fused to a DNA-binding domain⁴⁹.

TGF- β -induced transcriptional activation versus repression

TGF- β proteins activate transcription through physical interaction and functional cooperation of DNA-binding Smads with sequence-specific transcription factors and the coactivators CBP and p300 (Fig. 5). R-Smads (except Smad2) and Smad4 bind to preferred DNA sequences with a 100-fold lower affinity than the interacting high-affinity DNA-binding transcription factors, yet their DNA binding is required for transcriptional activation. Selective DNA binding to a subset of promoters that bind a potential Smad-interacting transcription factor defines the promoters that are activated in response to the ligand. The number of DNA-binding transcription factors with which Smads can functionally interact is impressive^{1–3}, and these are also often regulated by multiple signalling pathways. Besides the essential CBP or p300 coactivators, other coactivators and co-repressors that interact with Smads define the level of transcriptional activation. Smad4 itself acts as a key coactivator that enhances ligand-induced transcription by stabilizing the interaction of the R-Smads with DNA and CBP/p300. This model of transcriptional activation has been reviewed extensively elsewhere^{1–3} and will not be discussed here.

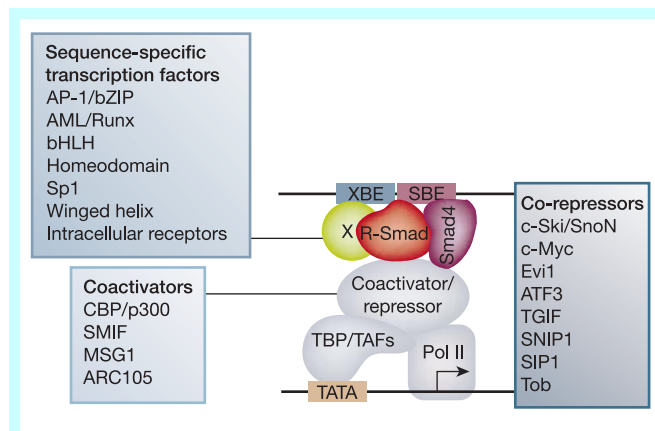


Figure 5 The R-Smad–Smad4 complex cooperates with sequence-specific transcription factors (X) that bind with high affinity to a cognate DNA sequence (XBE), yet also binds with lower affinity to a Smad-binding DNA element (SBE) to activate transcription in response to TGF- β ligand. R-Smads interact directly with the essential CBP or p300 coactivator, and Smad4 serves as coactivator for R-Smads by stabilizing the R-Smad interaction with CBP/p300. Other Smad-interacting coactivators, such as SMIF, MSG1, Swift and ARC105, further define the level of Smad-mediated transcription activation (see refs 1–3, 97, 98 and references therein). Smad-interacting co-repressors downregulate Smad-mediated transactivation. Several of these are proto-oncogenes—for example, c-Ski and the related SnoN, c-Myc^{99,100} and Evi1—linking malignant transformation to repression of TGF- β /Smad-induced transcription. Other Smad co-repressors—for example, the homeodomain proteins TGIF (TGF- β -induced factor) and SNIP1 (Smad nuclear interacting protein)—repress not only TGF- β /Smad-mediated transcriptional activation, but also Smad-independent transcription. Interaction of Tob with BMP-activated Smads represses BMP-activated gene expression, whereas its interaction with Smad2 represses interleukin-2 expression in T cells.

Many genes are activated in response to TGF- β ligands, whereas others are transcriptionally repressed. Smad co-repressors that inhibit transcriptional activation by Smads (Fig. 5) have not been implicated in TGF- β -induced repression, raising the question of what defines transcriptional activation versus repression by Smads.

TGF- β inhibits cell-cycle progression by regulating the transcription of cell-cycle regulators. Among them, c-Myc and Id family members are downregulated by TGF- β ^{60,61}. In cells with TGF- β -induced downregulation of c-Myc expression, Smad3 represses c-Myc transcription in association with the transcription factors E2F4 and E2F5, and the co-repressor p107. This complex is pre-assembled in the cytoplasm and, in response to TGF- β treatment, translocates into the nucleus, where, in association with Smad4, it binds to a Smad–E2F-binding site in the c-Myc promoter and represses c-Myc expression⁶⁰. In Id1 downregulation, TGF- β -activated Smad3 directly induces ATF3 expression, and ATF3 and Smad3 then form a complex that represses the Id1 promoter⁶¹. Similarly, SIP1, induced by TGF- β , downregulates E-cadherin expression⁶².

Apart from the repression of transcription of Id members in several differentiation lineages, TGF- β also inhibits myoblast, osteoblast and adipocyte differentiation through functional repression of key transcription factors that drive these differentiation pathways. Smad3 represses transcription by Runx2/CBFA1 (core-binding factor, runt domain, α -subunit) in osteoblastic differentiation⁶³, MyoD and other myogenic basic helix–loop–helix (bHLH) transcription factors in myoblasts⁶⁴, and CCAAT/enhancer-binding proteins (C/EBPs) in adipocyte differentiation⁶⁵. Smad3 represses MyoD function by physically interacting with the HLH domain of MyoD, obstructing MyoD function dimerization with E12/E47, which is required for efficient DNA binding to E-box sequences⁶⁴. Smad3 does not interfere with DNA binding of Runx2 or C/EBPs, and represses the transcription function of C/EBP^{63,65}. Whether Smads repress or activate transcription also depends on the cell type and the promoter sequence. Smad3 cooperates with Runx proteins to activate transcription in epithelial cells, and represses transcription from the same promoter in mesenchymal cells. The Runx-recognition-sequence context defines Smad-mediated activation versus repression⁶³. Finally, the MH1 domain of Smad3 can interact with histone deacetylases⁶⁶. Thus, distinct mechanisms, depending on the interacting transcription factor, the promoter and the intracellular context lead to Smad-mediated repression and determine whether Smads activate or repress transcription.

Regulation through receptor proteins

TGF- β receptor complexes are nodal points for multiprotein assemblies that regulate receptor function, routing, and Smad and non-Smad signalling pathways. These interactions are likely to depend on the activation and subcellular localization of the receptors and vary throughout the 'life cycle' of the receptor complex.

In addition to anchoring proteins that link Smads to their receptors (Fig. 4), several WD-40 repeat proteins can associate with TGF- β receptors (Fig. 6). For example, TGF- β -receptor-interacting protein 1 (TRIP-1) interacts with the ligand-bound TGF- β type II receptor and is phosphorylated by it. Increased TRIP-1 levels repress TGF- β -induced transcription, and some mutants enhance TGF- β responses, presumably through interference with endogenous TRIP-1 (ref. 67). TRIP-1 also participates in the translation-initiation factor complex eIF3, possibly linking TRIP-1 to TGF- β -induced translational regulation. Such a linkage is also suggested for the translation factor eIF2 α , which, similarly to TRIP-1, interacts with the type II receptor and downregulates the gene-expression response⁶⁸. The regulatory B α subunit of protein phosphatase 2A (PP2A), another WD-40 repeat protein, can also directly interact with type I TGF- β receptors, and enhances the anti-mitogenic signalling of TGF- β ⁶⁹. Upon receptor activation, PP2A-B α binds

to type I receptors, possibly activating PP2A-mediated signalling (see below). A third WD-40 protein, STRAP, can associate with the type I and type II TGF- β receptors, and with Smad2, Smad3 and Smad7. STRAP stabilizes the Smad7 association with the receptor and inhibits TGF- β -dependent transcription, probably by interfering with Smad2 or Smad3 binding to the receptor⁷⁰.

TGF- β receptors also interact with the immunophilin FKBP12 at a conserved Leu-Pro motif adjacent to the GS domain of unliganded type I receptors. FKBP12 decreases TGF- β signalling by inhibiting type I receptor phosphorylation by the type II receptor⁷¹, which correlates with decreased receptor internalization and Smad phosphorylation⁷². Type I receptors defective in FKBP12 binding have elevated basal signalling, but normal signalling in response to TGF- β , and inactivation of FKBP12 expression produces a phenotype consistent with TGF- β receptor hyperactivity⁷³. Thus, FKBP12 may act as a safeguard against leaky TGF- β receptor signalling. Although FKBP12 can have a role in calcineurin and NF-AT signalling, there is no evidence that TGF- β induces FKBP12-dependent signalling. Also the α -subunit of farnesyl transferase can interact with the type I receptor, but no response has been correlated with this interaction⁷⁴.

Smad-independent signalling through MAPK pathways

Besides Smad-mediated transcription, TGF- β activates other signalling cascades, including MAPK pathways (Fig. 6). Some of these pathways regulate Smad activation, as described above, but others might induce responses unrelated to transcription.

TGF- β can activate the Erk, JNK and p38 MAPK kinase pathways. Activation with slow kinetics in some cases may result from Smad-dependent transcription responses, but the rapid activation (5–15 min) in other cases suggests independence from transcription¹. Studies using Smad4-deficient cells, or dominant-negative Smads, support the possibility of MAPK pathway activation that is independent from Smads⁵⁴. In addition, mutated TGF- β type I receptors, defective in Smad activation, activate p38 MAPK signalling in response to TGF- β ⁷⁵.

The mechanisms of Erk, JNK or p38 MAPK activation by TGF- β and its biological consequences are poorly characterized. Rapid

activation of Ras by TGF- β in epithelial cells may implicate Ras in TGF- β -induced Erk MAPK signalling⁷⁶. JNK and p38 MAPK signalling are activated by various MAPK kinase kinases (MAPKKKs) in response to many stimuli. Both TGF- β and BMP-4 can activate TGF- β -activated kinase 1 (TAK1), a MAPKKK family member⁷⁷. Perhaps XIAP (X-linked inhibitor of apoptosis) links TGF- β /BMP receptor activation to TAK1 signalling⁷⁷; however, a direct interaction between XIAP and type I receptors has not been demonstrated, and XIAP-deficient mice respond to TGF- β ⁷⁸. MEKK1 may also function upstream of TGF- β -mediated activation of MAPKKs; thus, MEKK1 and TAK1 could activate JNK through MAPK kinase 4 (MKK4), and p38 MAPK through MKK3 or MKK6, in response to TGF- β . Because TAK1 can phosphorylate and activate I κ B kinase, thus stimulating NF- κ B signalling, TGF- β /BMP signalling may induce NF- κ B signalling. Further characterization of this web of interactions will provide insight into the activation of MAPK pathways by TGF- β ligands.

TGF- β -induced activation of the Erk and JNK pathways can result in Smad phosphorylation and regulate Smad activation^{51–54}. Also, TGF- β -induced activation of Ras/Erk MAPK signalling can induce TGF- β 1 expression, thereby amplifying the TGF- β response and inducing secondary TGF- β responses⁷⁶. Activation of MAPK pathways by TGF- β may also affect transcription responses through direct effects on Smad-interacting transcription factors—for example, the JNK substrate c-Jun or the p38 MAPK substrate ATF-2 (activating transcription factor 2)—allowing convergence of TGF- β -induced Smad and MAPK pathways^{1–3}. The dual ability of TGF- β to activate Smads and MAPK signalling has a role in TGF- β -induced epithelial-to-mesenchymal transdifferentiation, which depends in part on the Erk and/or p38 MAPK pathways^{75,79,80}. Although this convergence often results in cooperativity, these pathways may also counteract each other. For example, Smad6 can bind to TAK1 and downregulate its activity⁸¹, whereas Smad7 can enhance and sustain JNK activation⁸². Also, c-Jun inhibits Smad2 signalling through association with Smad co-repressors, and this interaction is regulated by JNK signalling⁸³. Thus, the balance between direct activation of Smads and MAPK pathways often defines cellular responses to TGF- β .

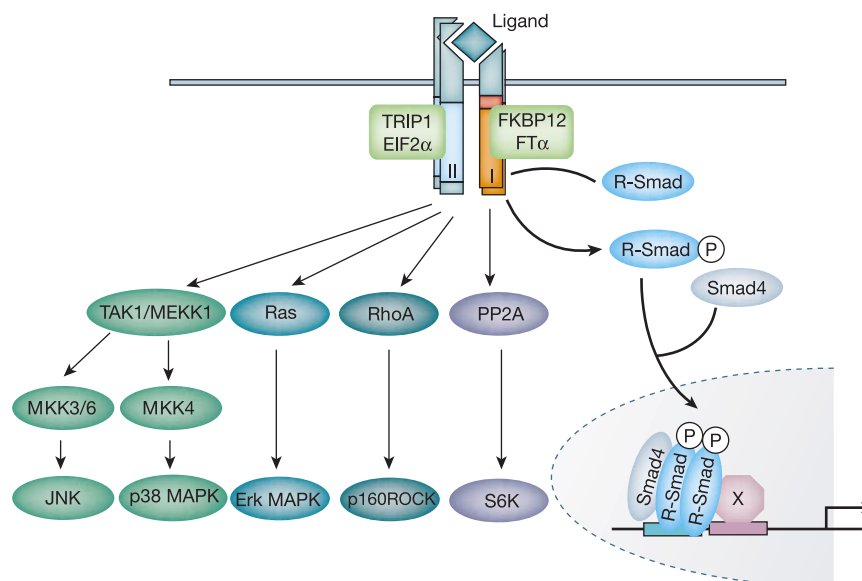


Figure 6 TGF- β receptor signalling through Smad-independent pathways. Apart from proteins that interact with receptors and Smads (see Fig. 4), other proteins (green) can associate with the type II or type I receptors and regulate TGF- β ligand signalling

without an apparent direct effect on Smad activation. In addition, the activated receptor complex activates non-Smad signalling pathways, such as MAPK, PP2A/p70^{S6K}, RhoA and TAK1/MEKK1. Only the best-characterized pathways are shown.

Other TGF- β -induced signalling pathways

Depending on the cell line, TGF- β can rapidly activate Rho-like GTPases (Fig. 6), including RhoA, Rac and Cdc42, although delayed activation of RhoA and Cdc42, because of new protein synthesis, has also been observed^{79,84,85}. TGF- β also enhances the expression of RhoB⁸⁶, possibly by inhibiting its proteasomal degradation, and induces the expression of NET1, a RhoA-specific guanine exchange factor that mediates RhoA activation⁸⁷. Finally, Ras activation in response to TGF- β may also lead to activation of a Rho-like GTPase.

Rac and Cdc42 regulate JNK and p38 MAPK pathway activation, presumably by directly interacting with MAPKKs upstream of JNK and p38 MAPK, whereas Rho, Rac and Cdc42 affect the cytoskeletal organization⁸⁸. Activation of small GTPases by TGF- β may play a part in gene-expression responses. RhoB counteracts⁸⁶, whereas Rac1 contributes to⁷⁹, TGF- β -induced gene expression, and Rho-dependent activation of JNK signalling contributes to Smad activation⁸⁴. These GTPases also mediate TGF- β -induced changes in cytoskeletal organization and epithelial-to-mesenchymal transdifferentiation. Activation of Rac1, RhoA and p38 MAPK, an effector of Cdc42, is required for rapid membrane ruffling and lamellipodia formation in response to TGF- β ⁸⁵. Activation of RhoA and its effector kinase p160^{ROCK}, as well as Cdc24, p38 MAPK and Smad signalling have been implicated in TGF- β -induced stress-fibre formation and epithelial-to-mesenchymal transdifferentiation^{75,79,84,85,87}.

TGF- β can also activate phosphatidylinositol-3-kinase (PI3K), as indicated by phosphorylation of its effector Akt^{89,90}. This activation can be direct, with possible involvement of RhoA⁸⁹, but can also result from TGF- β -induced TGF- α expression and consequent EGF receptor activation⁹⁰. Chemical inhibition of PI3K activity reduces TGF- β -induced Smad2 phosphorylation and transcription, and epithelial-to-mesenchymal transdifferentiation⁸⁹; however, inhibition of PI3K at lower inhibitor concentrations did not affect TGF- β -induced epithelial-to-mesenchymal differentiation^{7,91}. The role of PI3K in differentiation remains to be clarified.

TGF- β may also signal through PP2A, which consists of a catalytic C subunit and a regulatory A subunit, with which a regulatory B subunit can interact. Upon TGF- β binding, the B α subunit of PP2A associates with the activated type I receptor⁶⁹. Because B α inhibits the activity of the A/C dimer, this recruitment is expected to enhance PP2A activity. Upon TGF- β stimulation, the A and C subunits, as well as B α , interact with p70^{S6K}, a kinase with a key role in translational control and cell-cycle progression⁹². The resulting dephosphorylation and decreased activity of p70^{S6K} is thought to contribute to TGF- β -induced growth arrest independently of Smads⁹², suggesting a role for PP2A in the response to TGF- β . Protein phosphatase 1 (PP1) has also been implicated in TGF- β signalling. Its catalytic subunit (PP1c) interacts with *Drosophila* SARA, and negatively regulates Dpp signalling by affecting type I receptor phosphorylation⁹³.

Finally, TGF- β signalling also regulates protein stability. Most notably, TGF- β increases the degradation of the Smad co-repressor SnoN^{24,94,95} and the T β RI receptor^{36,37}. TGF- β -induced degradation of SnoN involves Smad2 or Smad3 as docking proteins to target it for proteasomal degradation. Thus, Smad2/3 interact with SnoN and Smurf2 or the anaphase-promoting complex, which then serve as E3 ubiquitin ligases for SnoN^{19,94,95}. Similarly, in response to TGF- β , Smad7 interacts with Smurf1 or Smurf2, and interaction of this complex with T β RI leads to receptor ubiquitination and degradation^{36,37}. It remains to be determined whether other Smad-interacting proteins are subject to TGF- β -induced degradation, and whether ubiquitin-mediated degradation contributes to the cellular response to TGF- β family signalling.

Future perspectives

Extensive progress has provided insight into the complex regulation and roles of Smads in transcription responses, and the effects of

signalling crosstalk on Smad function. The demonstration of Smad-independent TGF- β signalling pathways and the regulation of TGF- β signalling by receptor-associated proteins illustrate that our understanding now transcends the linear model of Smad signalling as an effector of TGF- β signalling. Incorporation of genomic and proteomic approaches, combined with RNA interference technology and mouse genetic manipulation, will advance our understanding of the plasticity of the cellular response to TGF- β , and the roles of TGF- β signalling components in coordinating signalling events and maintaining cell and tissue homeostasis.

Note added in proof: A recent study¹⁰¹ demonstrated direct interaction of the cytoplasmic domain of BMP-RII with the cytoskeletal regulator LIM kinase 1 as well as BMP-induced regulation of the kinase activity of LIM kinase 1. These data provide evidence for direct signalling by a type II receptor, through LIM kinase 1, to the cytoskeleton. □

doi:10.1038/nature02006.

1. Massagué, J. How cells read TGF- β signals. *Nature Rev. Mol. Cell Biol.* **1**, 169–178 (2000).
2. Itoh, S., Itoh, F., Goumans, M. J. & ten Dijke, P. Signaling of transforming growth factor- β family members through Smad proteins. *Eur. J. Biochem.* **267**, 6954–6967 (2000).
3. Moustakas, A., Souchevsky, S. & Heldin, C.-H. Smad regulation in TGF- β signal transduction. *J. Cell Sci.* **114**, 4359–4369 (2001).
4. Derynck, R. & Feng, X.-H. TGF- β receptor signaling. *Biochim. Biophys. Acta* **1333**, F105–F150 (1997).
5. Gilboa, L. *et al.* Bone morphogenetic protein receptor complexes on the surface of live cells: a new oligomerization mode for serine/threonine kinase receptors. *Mol. Biol. Cell* **11**, 1023–1035 (2000).
6. Goumans, M. J. *et al.* Balancing the activation state of the endothelium via two distinct TGF- β type I receptors. *EMBO J.* **21**, 1743–1753 (2002).
7. Miettinen, P. J., Ebner, R., Lopez, A. R. & Derynck, R. TGF- β induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. *J. Cell Biol.* **127**, 2021–2036 (1994).
8. Lai, Y. T. *et al.* Activin receptor-like kinase 2 can mediate atrioventricular cushion transformation. *Dev. Biol.* **222**, 1–11 (2000).
9. Yan, Y. T. *et al.* Dual roles of Cripto as a ligand and coreceptor in the nodal signaling pathway. *Mol. Cell Biol.* **22**, 4439–4449 (2002).
10. Blobe, G. C., Liu, X., Fang, S. J., How, T. & Lodish, H. F. A novel mechanism for regulating transforming growth factor β (TGF- β) signaling. Functional modulation of type III TGF- β receptor expression through interaction with the PDZ domain protein, GIPC. *J. Biol. Chem.* **276**, 39608–39617 (2001).
11. Barbara, N. P., Wrana, J. L. & Letarte, M. Endoglin is an accessory protein that interacts with the signaling receptor complex of multiple members of the transforming growth factor- β superfamily. *J. Biol. Chem.* **274**, 584–594 (1999).
12. Masuyama, N., Hanafusa, H., Kusabe, M., Shibuya, H. & Nishida, E. Identification of two Smad4 proteins in *Xenopus*. Their common and distinct properties. *J. Biol. Chem.* **274**, 12163–12170 (1999).
13. Sirard, C. *et al.* Targeted disruption in murine cells reveals variable requirement for Smad4 in transforming growth factor β -related signaling. *J. Biol. Chem.* **275**, 2063–2070 (2000).
14. Durocher, D. *et al.* The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell* **6**, 1169–1182 (2000).
15. Choy, L., Skillington, J. & Derynck, R. Roles of autocrine TGF- β receptor and Smad signaling in adipocyte differentiation. *J. Cell Biol.* **149**, 667–682 (2000).
16. Arora, K. & Warrior, R. A new Smurf in the village. *Dev. Cell* **1**, 441–442 (2001).
17. Zhu, H., Kavasaki, P., Abdollah, S., Wrana, J. L. & Thomsen, G. H. A SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation. *Nature* **400**, 687–693 (1999).
18. Zhang, Y., Chang, C., Gehring, D. J., Hemmati-Brivanlou, A. & Derynck, R. Regulation of Smad degradation and activity by Smurf2, an E3 ubiquitin ligase. *Proc. Natl Acad. Sci. USA* **98**, 974–979 (2001).
19. Bonni, S. *et al.* TGF- β induces assembly of a Smad2-Smurf2 ubiquitin ligase complex that targets SnoN for degradation. *Nature Cell Biol.* **3**, 587–595 (2001).
20. Lo, R. S. & Massagué, J. Ubiquitin-dependent degradation of TGF- β -activated Smad2. *Nature Cell Biol.* **1**, 472–478 (1999).
21. Fukuchi, M. *et al.* Ligand-dependent degradation of Smad3 by a ubiquitin ligase complex of ROC1 and associated proteins. *Mol. Biol. Cell* **12**, 1431–1443 (2001).
22. Inman, G. J., Nicolas, F. J. & Hill, C. S. Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF- β receptor activity. *Mol. Cell* **10**, 283–294 (2002).
23. Xu, L., Kang, Y., Col, S. & Massagué, J. Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGF- β signaling complexes in the cytoplasm and nucleus. *Mol. Cell* **10**, 271–282 (2002).
24. Lee, P. S., Chang, C., Liu, D. & Derynck, R. Sumoylation of Smad4, the common Smad mediator of TGF- β family signaling. *J. Biol. Chem.* **278**, 27853–27863 (2003).
25. Xu, J. & Attisano, L. Mutations in the tumor suppressors Smad2 and Smad4 inactivate transforming growth factor β signaling by targeting Smads to the ubiquitin-proteasome pathway. *Proc. Natl Acad. Sci. USA* **97**, 4820–4825 (2000).
26. Wan, M. *et al.* Jab1 antagonizes TGF- β signaling by inducing Smad4 degradation. *EMBO Rep.* **3**, 171–176 (2002).
27. Huse, M. *et al.* The TGF- β receptor activation process: an inhibitor- to substrate-binding switch. *Mol. Cell* **8**, 671–682 (2001).
28. Penheiter, S. G. *et al.* Internalization-dependent and -independent requirements for transforming growth factor β receptor signaling via the Smad pathway. *Mol. Cell Biol.* **22**, 4750–4759 (2002).

29. Hayes, S., Chawla, A. & Corvera, S. TGF- β receptor internalization into EEA1-enriched early endosomes: role in signaling to Smad2. *J. Cell Biol.* **158**, 1239–1249 (2002).
30. Di Guglielmo, G. M., Le Roy, C., Goodfellow, A. F. & Wrana, J. L. Distinct endocytic pathways regulate TGF- β receptor signaling and turnover. *Nature Cell Biol.* **5**, 410–421 (2003).
31. Hocevar, B. A., Smine, A., Xu, X. X. & Howe, P. H. The adaptor molecule Disabled-2 links the transforming growth factor β receptors to the Smad pathway. *EMBO J.* **20**, 2789–2801 (2001).
32. Razani, B. *et al.* Caveolin-1 regulates transforming growth factor (TGF)- β /SMAD signaling through an interaction with the TGF- β type I receptor. *J. Biol. Chem.* **276**, 6727–6738 (2001).
33. Dong, C., Li, Z., Alvarez, R. Jr, Feng, X.-H. & Goldschmidt-Clermont, P. J. Microtubule binding to Smads may regulate TGF- β activity. *Mol. Cell* **5**, 27–34 (2000).
34. Sasaki, A., Masuda, Y., Ohta, Y., Ikeda, K. & Watanabe, K. Filamin associates with Smads and regulates transforming growth factor- β signaling. *J. Biol. Chem.* **276**, 17871–17877 (2001).
35. Tang, Y. *et al.* Disruption of transforming growth factor- β signaling in ELF- β -spectrin-deficient mice. *Science* **299**, 574–577 (2003).
36. Kavask, P. *et al.* Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF- β receptor for degradation. *Mol. Cell* **6**, 1365–1375 (2000).
37. Ebisawa, T. *et al.* Smurf1 interacts with transforming growth factor- β type I receptor through Smad7 and induces receptor degradation. *J. Biol. Chem.* **276**, 12477–12480 (2001).
38. Wu, J. W. *et al.* Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF- β signaling. *Mol. Cell* **8**, 1277–1289 (2001).
39. Chacko, B. M. *et al.* The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization. *Nature Struct. Biol.* **8**, 248–253 (2001).
40. Inman, G. J. & Hill, C. S. Stoichiometry of active Smad-transcription factor complexes on DNA. *J. Biol. Chem.* **277**, 51008–51016 (2002).
41. Feng, X.-H., Lin, X. & Derynck, R. Smad2 Smad3 and Smad4 cooperate with Sp1 to induce p15^{INK4B} transcription in response to TGF- β . *EMBO J.* **19**, 5178–5193 (2000).
42. Watanabe, M., Masuyama, N., Fukuda, M. & Nishida, E. Regulation of intracellular dynamics of Smad4 by its leucine-rich nuclear export signal. *EMBO Rep.* **1**, 176–182 (2000).
43. Xiao, Z., Watson, N., Rodriguez, C. & Lodish, H. F. Nucleocytoplasmic shuttling of Smad1 conferred by its nuclear localization and nuclear export signals. *J. Biol. Chem.* **276**, 39404–39410 (2001).
44. Xu, L., Chen, Y.-G. & Massagué, J. The nuclear import function of Smad2 is masked by SARA and unmasked by TGF β -dependent phosphorylation. *Nature Cell Biol.* **2**, 559–562 (2000).
45. Kurisaki, A., Kose, S., Yoneda, Y., Heldin, C. H. & Moustakas, A. Transforming growth factor- β induces nuclear import of Smad3 in an importin- β 1 and Ran-dependent manner. *Mol. Biol. Cell* **12**, 1079–1091 (2001).
46. Itoh, F. *et al.* Promoting bone morphogenetic protein signaling through negative regulation of inhibitory Smads. *EMBO J.* **20**, 4132–4142 (2001).
47. Hanyu, A. *et al.* The N domain of Smad7 is essential for specific inhibition of transforming growth factor- β signaling. *J. Cell Biol.* **155**, 1017–1027 (2001).
48. Bai, S. & Cao, X. A nuclear antagonistic mechanism of inhibitory Smads in transforming growth factor- β signaling. *J. Biol. Chem.* **277**, 4176–4182 (2002).
49. Pulaski, L., Landström, M., Heldin, C. H. & Souchevlynskyi, S. Phosphorylation of Smad7 at Ser-249 does not interfere with its inhibitory role in transforming growth factor- β -dependent signaling but affects Smad7-dependent transcriptional activation. *J. Biol. Chem.* **276**, 14344–14349 (2001).
50. Grönroos, E., Hellman, U., Heldin, C. H. & Ericsson, J. Control of Smad7 stability by Competition between acetylation and ubiquitination. *Mol. Cell* **10**, 483–493 (2002).
51. de Caestecker, M. P. *et al.* Smad2 transduces common signals from receptor serine-threonine and tyrosine kinases. *Genes Dev.* **12**, 1587–1592 (1998).
52. Kretschmar, M., Doody, J., Timokhina, I. & Massagué, J. A mechanism of repression of TGF- β /Smad signaling by oncogenic Ras. *Genes Dev.* **13**, 804–816 (1999).
53. Funaba, M., Zimmerman, C. M. & Mathews, L. S. Modulation of Smad2-mediated signaling by extracellular signal-regulated kinase. *J. Biol. Chem.* **277**, 41361–41368 (2002).
54. Engel, M. E., McDonnell, M. A., Law, B. K. & Moses, H. L. Interdependent SMAD and JNK signaling in transforming growth factor- β -mediated transcription. *J. Biol. Chem.* **274**, 37413–37420 (1999).
55. Janda, E. *et al.* Ras and TGF- β cooperatively regulate epithelial cell plasticity and metastasis: dissection of Ras signaling pathways. *J. Cell Biol.* **156**, 299–313 (2002).
56. Grimm, O. H. & Gurdon, J. B. Nuclear exclusion of Smad2 is a mechanism leading to loss of competence. *Nature Cell Biol.* **4**, 519–522 (2002).
57. Brown, J. D., DiChiara, M. R., Anderson, K. R., Gimbrone, M. A. Jr & Topper, J. N. MEKK-1, a component of the stress (stress-activated protein kinase/c-Jun N-terminal kinase) pathway, can selectively activate Smad2-mediated transcriptional activation in endothelial cells. *J. Biol. Chem.* **274**, 8797–8805 (1999).
58. Wicks, S. J., Lui, S., Abdel-Wahab, N., Mason, R. M. & Chantry, A. Inactivation of smad-transforming growth factor β signaling by Ca²⁺-calmodulin-dependent protein kinase II. *Mol. Cell Biol.* **20**, 8103–8111 (2000).
59. Yakymovych, I., ten Dijke, P., Heldin, C. H. & Souchevlynskyi, S. Regulation of Smad signaling by protein kinase C. *FASEB J.* **15**, 553–555 (2001).
60. Chen, C.-R., Kang, Y., Siegel, P. M. & Massagué, J. E2F4/5 and p107 as Smad cofactors linking the TGF β receptor to c-myc repression. *Cell* **110**, 19–32 (2002).
61. Kang, Y., Chen, C.-R. & Massagué, J. A self-enabling TGF- β response coupled to stress signaling: Smad engages stress response factor ATF3 for *Id1* repression in epithelial cells. *Mol. Cell* **11**, 915–926 (2003).
62. Comijn, J. *et al.* The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol. Cell* **7**, 1267–1278 (2001).
63. Alliston, T., Choy, L., Ducy, P., Karsenty, G. & Derynck, R. TGF- β -induced repression of CBFA1 by Smad3 decreases cbfa1 and osteocalcin expression and inhibits osteoblast differentiation. *EMBO J.* **20**, 2254–2272 (2001).
64. Liu, D., Black, B. L. & Derynck, R. TGF- β inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3. *Genes Dev.* **15**, 2950–2966 (2001).
65. Choy, L. & Derynck, R. Transforming growth factor- β inhibits adipocyte differentiation by Smad3 interacting with CCAAT/enhancer-binding protein (C/EBP) and repressing C/EBP transactivation function. *J. Biol. Chem.* **278**, 9609–9619 (2003).
66. Liberati, N. T., Moniwa, M., Borton, A. J., Davie, J. R. & Wang, X.-F. An essential role for Mad homology domain 1 in the association of Smad3 with histone deacetylase activity. *J. Biol. Chem.* **276**, 22595–22603 (2001).
67. Choy, L. & Derynck, R. The type II transforming growth factor (TGF)- β receptor-interacting protein TRIP-1 acts as a modulator of the TGF- β response. *J. Biol. Chem.* **273**, 31455–31462 (1998).
68. McGonigle, S., Beall, M. J. & Pearce, E. J. Eukaryotic initiation factor 2 α subunit associates with TGF- β receptors and 14-3-3 and acts as a modulator of the TGF- β response. *Biochemistry* **41**, 579–587 (2002).
69. Griswold-Prenner, I., Kamibayashi, C., Maruoka, E. M., Mumby, M. C. & Derynck, R. Physical and functional interactions between type I transforming growth factor β receptors and B α , a WD-40 repeat subunit of phosphatase 2A. *Mol. Cell Biol.* **18**, 6595–6604 (1998).
70. Datta, P. K. & Moses, H. L. STRAP and Smad7 synergize in the inhibition of transforming growth factor β signaling. *Mol. Cell Biol.* **20**, 3157–3167 (2000).
71. Huse, M., Chen, Y.-G., Massagué, J. & Kuriyan, J. Crystal structure of the cytoplasmic domain of the type I TGF- β receptor in complex with FKBP12. *Cell* **96**, 425–436 (1999).
72. Yao, D., Doré, J. J. Jr & Leof, E. B. FKBP12 is a negative regulator of transforming growth factor- β receptor internalization. *J. Biol. Chem.* **275**, 13149–13154 (2000).
73. Aghdasi, B. *et al.* FKBP12, the 12-kDa FK506-binding protein, is a physiologic regulator of the cell cycle. *Proc. Natl Acad. Sci. USA* **98**, 2425–2430 (2001).
74. Ventura, F., Liu, F., Doody, J. & Massagué, J. Interaction of transforming growth factor- β receptor I with farnesyl-protein transferase- α in yeast and mammalian cells. *J. Biol. Chem.* **271**, 13931–13934 (1996).
75. Yu, L., Hebert, M. C. & Zhang, Y. E. TGF- β receptor-activated p38 MAP kinase mediates Smad-independent TGF- β responses. *EMBO J.* **21**, 3749–3759 (2002).
76. Yue, J. & Mulder, K. M. Activation of the mitogen-activated protein kinase pathway by transforming growth factor- β . *Methods Mol. Biol.* **142**, 125–131 (2000).
77. Yamaguchi, K. *et al.* XIAP, a cellular member of the inhibitor of apoptosis protein family, links the receptors to TAB1-TAK1 in the BMP signaling pathway. *EMBO J.* **18**, 179–187 (1999).
78. Salvendy, G. S. & Duckett, C. S. IAP proteins: blocking the road to death's door. *Nature Rev. Mol. Cell Biol.* **3**, 401–410 (2002).
79. Bakin, A. V., Rinehart, C., Tomlinson, A. K. & Arteaga, C. L. p38 mitogen-activated protein kinase is required for TGF β -mediated fibroblastic transdifferentiation and cell migration. *J. Cell Sci.* **115**, 3193–3206 (2002).
80. Zavadil, J. *et al.* Genetic programs of epithelial cell plasticity directed by transforming growth factor- β . *Proc. Natl Acad. Sci. USA* **98**, 6686–6691 (2001).
81. Kimura, N., Matsuo, R., Shibuya, H., Nakashima, K. & Taga, T. BMP2-induced apoptosis is mediated by activation of the TAK1-p38 kinase pathway that is negatively regulated by Smad6. *J. Biol. Chem.* **275**, 17647–17652 (2000).
82. Mazars, A. *et al.* Evidence for a role of the JNK cascade in Smad7-mediated apoptosis. *J. Biol. Chem.* **276**, 36797–36803 (2001).
83. Pessah, M. *et al.* c-Jun associates with the oncoprotein Ski and suppresses Smad2 transcriptional activity. *J. Biol. Chem.* **277**, 29094–29100 (2002).
84. Bhowmick, N. A. *et al.* Transforming growth factor- β 1 mediates epithelial to mesenchymal transdifferentiation through a RhoA-dependent mechanism. *Mol. Biol. Cell* **12**, 27–36 (2001).
85. Edlund, S., Landström, M., Heldin, C. H. & Aspenström, P. Transforming growth factor- β -induced mobilization of actin cytoskeleton requires signaling by small GTPases Cdc42 and RhoA. *Mol. Biol. Cell* **13**, 902–914 (2002).
86. Engel, M. E., Datta, P. K. & Moses, H. L. RhoB is stabilized by transforming growth factor β and antagonizes transcriptional activation. *J. Biol. Chem.* **273**, 9921–9926 (1998).
87. Shen, X. *et al.* The activity of guanine exchange factor NET1 is essential for transforming growth factor- β -mediated stress fiber formation. *J. Biol. Chem.* **276**, 15362–15368 (2001).
88. Bishop, A. L. & Hall, A. Rho GTPases and their effector proteins. *Biochem. J.* **348**, 241–255 (2000).
89. Bakin, A. V., Tomlinson, A. K., Bhowmick, N. A., Moses, H. L. & Arteaga, C. L. Phosphatidylinositol 3-kinase function is required for transforming growth factor- β -mediated epithelial to mesenchymal transition and cell migration. *J. Biol. Chem.* **275**, 36803–36810 (2000).
90. Vinals, F. & Pouyssegur, J. Transforming growth factor β 1 (TGF- β 1) promotes endothelial cell survival during *in vitro* angiogenesis via an autocrine mechanism implicating TGF- α signaling. *Mol. Cell Biol.* **21**, 7218–7230 (2001).
91. Piek, E., Moustakas, A., Kurisaki, A., Heldin, C. H. & ten Dijke, P. TGF- β type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells. *J. Cell Sci.* **112**, 4557–4568 (1999).
92. Petritsch, C., Beug, H., Balmain, A. & Ofi, M. TGF- β inhibits p70 S6 kinase via protein phosphatase 2A to induce G₁ arrest. *Genes Dev.* **14**, 3093–3101 (2000).
93. Bennett, D. & Alpheg, L. PP1 binds Sara and negatively regulates Dpp signalling in *Drosophila melanogaster*. *Nature Genet.* **31**, 419–423 (2002).
94. Stroschein, S. L., Bonni, S., Wrana, J. L. & Luo, K. Smad3 recruits the anaphase-promoting complex for ubiquitination and degradation of SnO. *Genes Dev.* **15**, 2822–2836 (2001).
95. Wan, Y., Liu, X. & Kirschner, M. W. The anaphase-promoting complex mediates TGF- β signaling by targeting SnO for destruction. *Mol. Cell* **8**, 1027–1039 (2001).
96. Yamakawa, N., Tsuchida, K. & Sugino, H. The rasGAP-binding protein, Dok-1, mediates activin signaling via serine/threonine kinase receptors. *EMBO J.* **21**, 1684–1694 (2002).
97. Bai, R. Y. *et al.* SMIF, a Smad4-interacting protein that functions as a co-activator in TGF β signalling. *Nature Cell Biol.* **4**, 181–190 (2002).
98. Kato, Y., Habas, R., Katsuyama, Y., Naar, A. M. & He, X. A component of the ARC/Mediator complex required for TGF- β /Nodal signalling. *Nature* **418**, 641–646 (2002).
99. Seoane, J. *et al.* TGF- β influences Myc, Miz-1 and Smad to control the CDK inhibitor p15^{INK4b}. *Nature Cell Biol.* **3**, 400–408 (2001).
100. Feng, X. H., Liang, Y. Y., Liang, M., Zhai, W. & Lin, X. Direct interaction of c-Myc with Smad2 and Smad3 to inhibit TGF- β -mediated induction of the CDK inhibitor p15^{INK4B}. *Mol. Cell* **9**, 133–143 (2002).
101. Foletta, V. C. *et al.* Direct signaling by the BMP type II receptor via the cytoskeletal regulator LIMK1. *J. Cell Biol.* **162**, 1089–1098 (2003).

Acknowledgements We apologize to the many researchers whose work could not be cited because of space limitations or was only cited indirectly by referring to reviews or more recent papers.

Correspondence and requests for materials should be addressed to R.D. (derynck@itsa.ucsf.edu) or Y.E.Z. (yingz@helix.nih.gov).

Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases

Simona Radutoiu¹, Lene Heegaard Madsen¹, Esben Bjørn Madsen¹, Hubert H. Felle², Yosuke Umehara^{1*}, Mette Grønlund¹, Shusei Sato³, Yasukazu Nakamura³, Satoshi Tabata³, Niels Sandal¹ & Jens Stougaard¹

¹Laboratory of Gene Expression, Department of Molecular Biology, University of Aarhus, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark

²Botanisches Institut I, Justus-Liebig Universität, D-35390 Giessen, Germany

³Kazusa DNA Research Institute, Kisarazu, Chiba 292-0812, Japan

* Present address: National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan

Although most higher plants establish a symbiosis with arbuscular mycorrhizal fungi, symbiotic nitrogen fixation with rhizobia is a salient feature of legumes. Despite this host range difference, mycorrhizal and rhizobial invasion shares a common plant-specified genetic programme controlling the early host interaction. One feature distinguishing legumes is their ability to perceive rhizobial-specific signal molecules. We describe here two LysM-type serine/threonine receptor kinase genes, *NFR1* and *NFR5*, enabling the model legume *Lotus japonicus* to recognize its bacterial microsymbiont *Mesorhizobium loti*. The extracellular domains of the two transmembrane kinases resemble LysM domains of peptidoglycan- and chitin-binding proteins, suggesting that they may be involved directly in perception of the rhizobial lipochitin-oligosaccharide signal. We show that *NFR1* and *NFR5* are required for the earliest physiological and cellular responses to this lipochitin-oligosaccharide signal, and demonstrate their role in the mechanism establishing susceptibility of the legume root for bacterial infection.

Legumes can acquire two important macronutrients—nitrogen and phosphorus—through symbiotic relationships with rhizobial bacteria and mycorrhizal fungi, respectively. Carbon from photosynthesis is exchanged for either nitrogen reduced by bacterial nitrogenase or phosphate extracted from the soil by the network of fungal hyphae. The host range of mycorrhizal fungi is broad, and the majority of all land plants form arbuscular mycorrhizal symbiosis^{1,2}. In contrast the host range of rhizobial symbiosis is narrow and, with one known exception, only legumes develop root nodules with rhizobia. Even within the legume family, specific and mutual recognition limits the host range of bacterial strains. Despite these differences, legume genetics connects fungal and rhizobial symbioses and provides evidence for the recruitment of components from a pre-existing plant–fungal pathway during evolution of the plant–bacterial interaction^{1,3}. For instance, several classes of legume mutants arrested early in rhizobial interaction are also impaired in mycorrhizal colonization^{4–6}. The genetic loci corresponding to these mutants define a common endosymbiotic pathway involved in legume–microsymbiont communication. In *Lotus* at least six genetic loci constitute the common pathway⁶, and the *SYMRK* gene was the first to be characterized molecularly. *SYMRK* encodes a leucine-rich repeat (LRR) receptor-like kinase that perceives both mycorrhizal and rhizobial signals, probably at the junction of the common pathway^{7,8}.

The nature of fungal signal molecules perceived by the plant root to allow hyphal entry and accommodation of the fungus is unknown⁹. In contrast, a combination of chemistry and bacterial genetics has identified lipochito-oligosaccharides (Nod-factors), consisting of substituted β ,1-4 *N*-acetylglucosamine (chitin) backbones, as the rhizobial morphogenic signal^{10,11} inducing root hair deformation and cell division leading to development of nodule primordia^{12–14}. The major Nod-factor secreted by the microsymbiont of *Lotus*, *M. loti*, is a pentameric *N*-acetylglucosamine, which carries a *cis*-vaccenic acid and a carbamoyl group at the non-reducing terminal residue together with a 4-*O*-acetylfucose at the reducing terminal residue^{13,14}. Several lines of evidence demonstrate that rhizobial host range is determined by the type and position of Nod-factor substituents, suggesting the involvement of one or more receptors¹⁵. For example, addition of acetyl-fucose to the *Rhizobium*

leguminosarum biovar *viciae* Nod-factor enables this rhizobial symbiont of pea to nodulate *Lotus*¹⁶. Further indications of receptor-mediated perception of Nod-factors come from physiological studies, where ion fluxes, membrane depolarization and Ca^{2+} spiking in legume root hairs is elicited within minutes of application of purified Nod-factor in the picomolar to nanomolar range^{17,18}.

Lotus plants carrying mutations in genes of the common pathway, such as *SYMRK*, show a Nod-factor-dependent root hair deformation response on rhizobial inoculation⁷, indicating that they are still able to sense the presence of bacterial signal(s) through upstream genes or a parallel pathway. Two mutant classes in *Lotus*—*nfr1* and *nfr5*—that have a wild-type mycorrhization phenotype^{4,19,20} but lack the root hair response to rhizobia, identify loci predicted to be involved in rhizobial signal perception or transduction. Here we show that *NFR1* and *NFR5* (ref. 21) are putative Nod-factor receptor genes necessary for Nod-factor perception upstream of *SYMRK* and the common pathway. Map-based cloning and characterization of *NFR1* and *NFR5* (ref. 21) reveals that both encode LysM-type serine/threonine receptor kinases.

NFR1 and *NFR5* mutant phenotypes

Structural requirements of Nod-factor molecules and the nanomolar concentrations eliciting root hair deformation and root cell divisions indicate that legume perception of the bacterial signal is receptor mediated. As a first step towards understanding the molecular mechanisms involved, we have characterized in detail the non-responsive phenotypes of the monogenic recessive *nfr1* and *nfr5* mutants (formerly known as *sym1* and *sym5*, respectively¹⁹). Three different strains of *M. loti*, NZP2235, R7A and TONO, all failed to induce infection threads or nodule primordia in mutant plants carrying the *nfr1-1*, *nfr1-2* or *nfr5-1*, *nfr5-2*, *nfr5-3* alleles. The earliest visible cellular change in wild-type plants is root hair deformation and root hair curling, easily observable 24 h after *M. loti* inoculation (Fig. 1a, b). These cellular responses were not observed in *nfr1* and *nfr5* mutants (Fig. 1c, d). The fact that lipochitin-oligosaccharides purified from the *M. loti* R7A strain also failed to induce any deformation of mutant root hairs suggests that both the *NFR1* and *NFR5* genes act early in the rhizobial interaction and are involved in Nod-factor perception (Fig. 1b, d).

NFR1 and NFR5 act upstream of the common pathway

SYMRK receptor kinase is required for both rhizobial and mycorrhizal invasion of *Lotus*, and has been suggested to be a key component of the receptor kinase complex, perceiving both bacterial and fungal signals through the extracellular LRR domain⁷. To determine whether SYMRK and NFR1 are in the same pathway their epistatic relationship was investigated by comparing the root hair deformation phenotype of *nfr1-2 symRK-3* double mutants to single mutant phenotypes. In *symRK* mutants, root hairs swell with balloon-shaped deformations in the presence of *M. loti* (Fig. 1e), but root hairs of *nfr1-2* mutants are unresponsive (Fig. 1c). Double mutants have no root hair response, indicating that NFR1-mediated Nod-factor perception is necessary for the Nod-factor response in *symRK* mutants (Fig. 1f). This interpretation is supported by the physiological response of root hairs towards purified Nod-factor (see below), where *nfr1-2 symRK-3* double mutants again respond similarly to *nfr1-2* mutants (Fig. 2c–e).

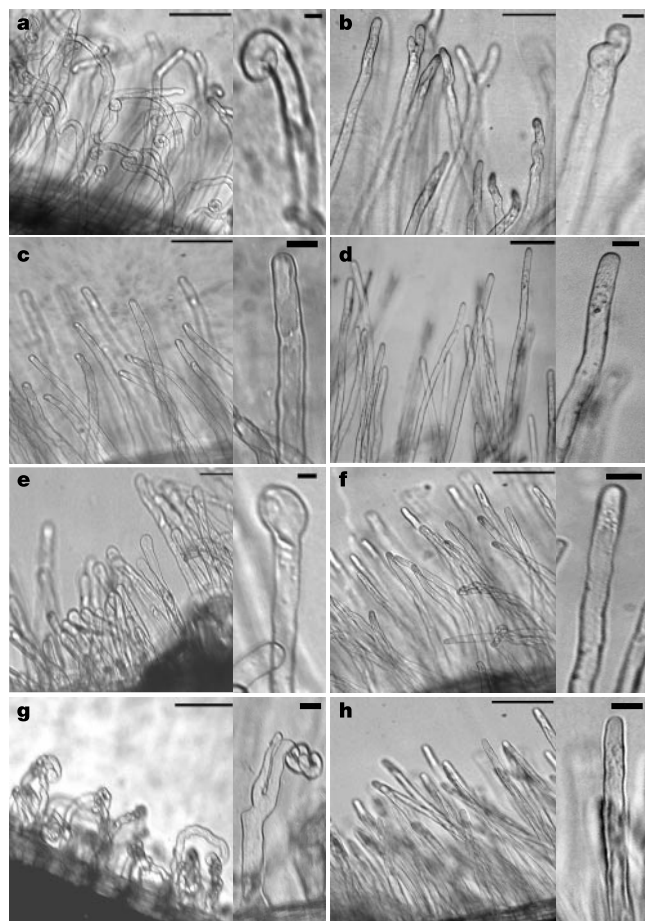


Figure 1 Root hair response after *M. loti* inoculation or Nod-factor application. **a**, Wild-type root hair curling on a seedling inoculated with *M. loti*. **b**, Wild-type root hair deformations after Nod-factor application. **c**, Root hairs on a *nfr1-1* seedling inoculated with *M. loti*. **d**, Root hairs on a *nfr1-1* seedling after Nod-factor application. **e**, Root hairs with balloon deformations on a *symRK-3* mutant inoculated with *M. loti*. **f**, Root hairs on a *nfr1-2 symRK-3* double mutant inoculated with *M. loti*. **g**, Excessive root hair response on *nin* mutants inoculated with *M. loti*. **h**, Root hairs on a *nfr1-2 nin* double mutant inoculated with *M. loti*. Root hairs on *nfr5-1* seedlings inoculated with *M. loti*, *nfr5-1* seedlings after Nod-factor application, together with untreated *nfr5-1*, wild-type, *nfr1-1* and *nfr1-2* controls, are indistinguishable from the straight root hairs shown in **c**, **d**, **f**, **h**, and are therefore not shown. Insets (right) show close-up views of root hairs. Scale bars: 50 μ m main images; insets 10 μ m.

Altogether, genetic and physiological results indicate that the NFR1 receptor kinase functions upstream and may precede SYMRK receptor kinase function in the rhizobial interaction.

Physiological experiments measuring ion fluxes in alfalfa (*Medicago sativa*) root hairs have shown that rapid calcium influx—a transient plasma membrane depolarization and an extracellular alkalinization of the root hair space—occurs within minutes of application of Nod-factor from *Sinorhizobium meliloti*^{17,18,22,23}. This implies that a specific receptor(s), present in epidermal root hair cells, binds the Nod-factor and transmits the signal(s). Similar to alfalfa, root hairs in *Lotus* undergo depolarization and extracellular alkalinization within minutes of exposure to *M. loti* Nod-factor (Fig. 2a), whereas mutants carrying *nfr5-1* and *nfr5-2* alleles have no detectable response (Fig. 2b). An attenuated and slower alkalinization was found in the *nfr1-1* mutants, whereas the *nfr1-2* mutants responded with acidification (Fig. 2c), reminiscent of the unspecific response of alfalfa to undecorated chitin octasaccharide molecules²⁴. Application of an undecorated chitin octasaccharide resulted in acidification of the root hair space in all mutants tested, indicating that similar to wild type, *nfr1* and *nfr5* mutants are capable of a general response to undecorated chitin octasaccharide molecules (Fig. 2f). Thus, both the NFR1 and NFR5 genes are essential for mounting the earliest detectable cellular and physiological responses to Nod-factor. Furthermore, the fact that *symRK-1* and *symRK-3* mutants, impaired in their symbiotic interactions with rhizobia and mycorrhiza, behave in a physiologically similar

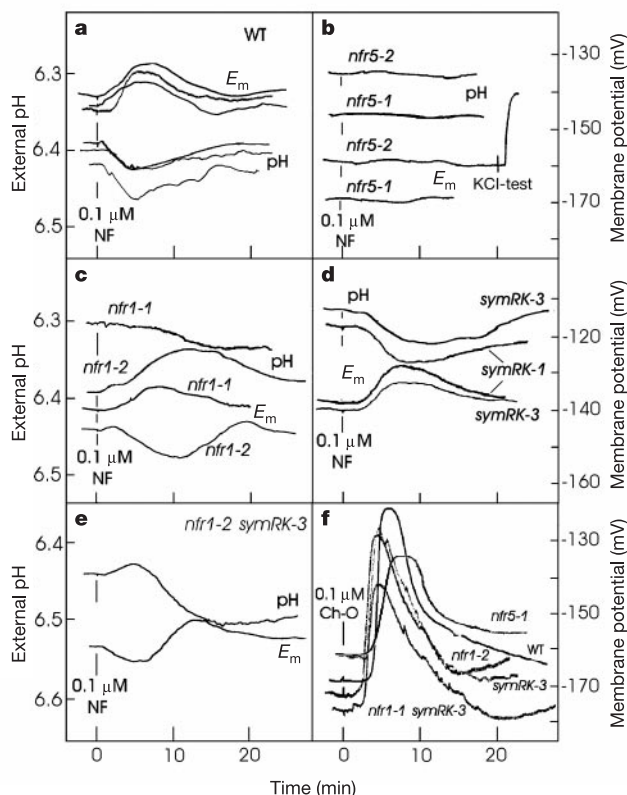


Figure 2 Membrane depolarization and pH changes in the extracellular root hair space after application of Nod-factor purified from *M. loti*. Graphs show the influence of 0.1 μ M Nod-factor (NF) on membrane potential (E_m) and/or external pH. **a**, Representative kinetics of three equivalent *Lotus* wild-type experiments. **b**, *nfr5-2* and *nfr5-1* mutants. **c**, *nfr1-1* and *nfr1-2* mutants. **d**, *symRK-1* and *symRK-3* mutants. **e**, *nfr1-2 symRK-3* double mutant. **f**, pH changes in the extracellular root hair space after application of an undecorated chitin octasaccharide.

way to wild type in response to purified Nod-factor implies that NFR1 and NFR5 function upstream of SYMRK and the common pathway (Fig. 2d, e). This interpretation is substantiated by an initial acidification response in the *nfr1-2 symRK-3* double mutant similar to the response of *nfr1-2* single mutants.

Expression of downstream symbiotic genes

The rhizobial-induced changes in root hair morphology and formation of the nodule primordia are accompanied by induction of specific plant symbiotic genes called early nodulins^{25,26}. Because *nfr1* and *nfr5* mutant plants failed to show root hair deformation and as the wild-type physiological response is lacking or attenuated, we followed the activation of two differentially regulated nodulin genes in *nfr1-1* and *nfr5-1* genotypes and, for comparison, in *symRK-1* (Fig. 3). One of the genes chosen for the analysis was *NIN*, which encodes a putative transcriptional regulator required for infection thread formation and inception of the nodule primordia²⁷. Normal mycorrhization and excessive root hair response in *nin* mutants places the function of this developmental regulator in the rhizobia-specific branch downstream of the common pathway⁶. The *ENOD2* gene encoding a hydroxyproline-rich protein expressed in nodule

parenchyma cells¹⁴ was chosen to represent early nodulins active during organogenesis. In wild-type plants, *NIN* transcripts were detectable 5 h after inoculation, and the transcript level continued to increase, whereas *ENOD2* expression peaked at day 12, at around the time that nodule cell differentiation is initiated (Fig. 3a, b). Purified Nod-factor applied to seedling roots also induced transcription of *NIN* in wild-type plants, where a 30-fold increase in steady-state levels of *NIN* messenger RNA 24 h after application was detected (Fig. 3c). By contrast, *NIN* and *ENOD2* transcripts were detected at levels comparable to untreated roots in *nfr1*, *nfr5* and *symRK-1* mutants on bacterial inoculation or Nod-factor application (Fig. 3a–c), demonstrating that functional *NFR1* and *NFR5* genes are required for activation of *NIN* and *ENOD2* during early signal perception and nodule organogenesis, respectively. The requirement for an upstream *NFR1* function was substantiated genetically by the absence of root hair deformation and curling in *nfr1-1 nin* double mutants (Fig. 1g, h).

Map-based cloning of *NFR1*

The *NFR1* gene was isolated using a positional cloning approach. On the genetic map of *Lotus* the *NFR1* locus is located on the short arm of chromosome I, approximately 22 cM from the top²⁸. Subsequent fine mapping in an F₂ population established from an ecotype (Miyakojima MG-20) cross and genotyping of 1,603 mutant plants identified markers delimiting the *NFR1* locus to 36 kilobases (kb) (Fig. 4a; see also Methods). A UFD1-like protein, a hypothetical protein and a receptor-like kinase were predicted to be encoded within this fragment (Fig. 4b). Owing to the predicted involvement of *NFR1* in signalling, the receptor-like kinase was a likely candidate gene and the corresponding gene region of the *nfr1-1* and *nfr1-2* mutant alleles was sequenced. Single nucleotide substitutions leading to premature stop codons were found in both the *nfr1-1* (CAA to TAA) and the *nfr1-2* (GAA to TAA) alleles. *In planta* complementation of the *nfr1* mutants was accomplished using the *Agrobacterium rhizogenes* transformed 'hairy roots' methodology²⁹. The wild-type *NFR1* gene including a 3,016 base pair (bp) promoter region and 3' region 392 bp downstream of the stop codon was introduced into *nfr1-1* and *nfr1-2* mutant plants through *A. rhizogenes*, and the nodulation phenotype was scored after inoculation with *M. loti*. Both *nfr1-1* and *nfr1-2* mutants were complemented with high efficiency (Table 1). Together the collective evidence obtained by genetic and physical mapping, sequencing of mutant alleles and successful complementation unequivocally identify the receptor-like kinase gene as *NFR1*. Two alternatively spliced complementary DNAs were identified using a combination of cDNA library screening and 5' rapid amplification of cDNA ends (RACE) on root RNA. Sequencing of nine 5' RACE products determined the transcription start site 115 bp in front of the start codon and the total length of the two cDNAs to 2,189 and 2,195 nucleotides, respectively. The 3' untranslated region (UTR) is 207 nucleotides (Fig. 4c). Alignment of genomic and cDNA sequences defined 12 exons in *NFR1* and a gene structure spanning 5,688 bp. Alternative splice donor sites at the 3' end of exon IV is responsible for the two alternative *NFR1* mRNAs of 2,189 and 2,195 nucleotides. Map-based cloning of the *NFR5* gene is described in the accompanying paper²¹.

The *NFR1* receptor kinase protein

The alternatively spliced *NFR1* cDNAs encode conceptual proteins of 621 and 623 amino acids, corresponding to 68.09 kDa and 68.23 kDa, respectively. The amino-terminal region consists of a signal peptide and two motifs with similarity to the LysM domain (Fig. 4d) (ref. 30). The carboxy terminus of the *NFR1* protein contains motifs associated with functional serine/threonine kinases^{31,32}. A transmembrane segment, predicted to anchor the *NFR1* receptor in the plasma membrane, is located between these domains. A topology with two extracellular LysM receiver domains

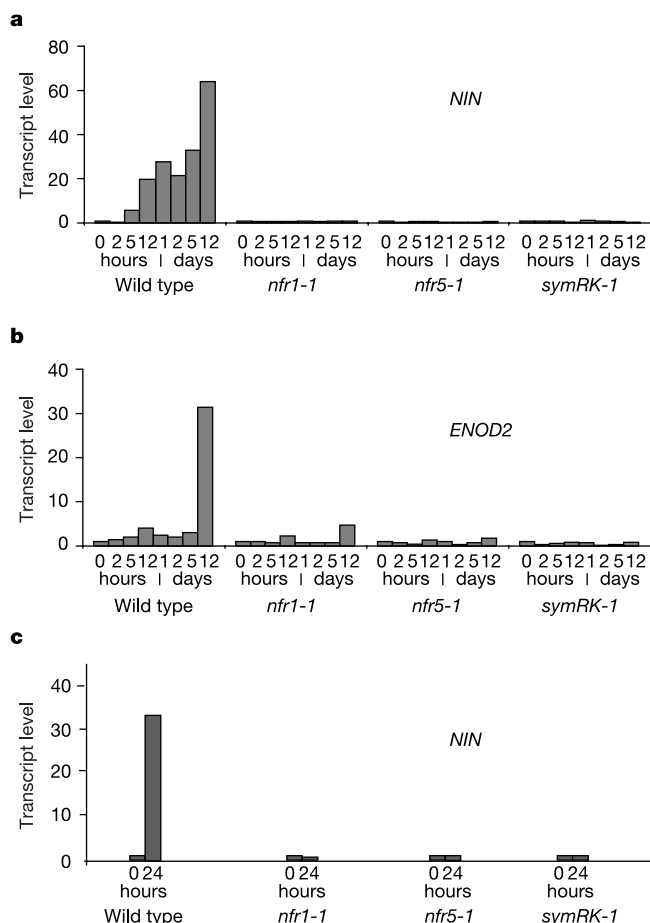


Figure 3 Expression of the *NIN* and *ENOD2* genes in wild type, *nfr1-1*, *nfr5-1* and *symRK-1* mutants. **a**, Steady-state *NIN* transcript levels in uninoculated roots and roots 2 h to 12 days after *M. loti* inoculation. **b**, *ENOD2* transcript levels in uninoculated roots and roots 2 h to 12 days after *M. loti* inoculation. **c**, Induction of *NIN* transcripts measured 24 h after application of purified Nod-factor. Transcript levels were measured by quantitative PCR and the identity of the amplified sequences was confirmed by sequencing. Results are shown as fold increase compared to uninoculated roots at time zero. ATPase was used as an internal control, and relative values normalized to the uninoculated root (0 h) are shown.

and an intracellular kinase is suggested by this analysis. In *nfr1-1* a stop codon is predicted to generate a translation product of 490/492 amino acids truncated in kinase domain VIII, and in *nfr1-2* a stop codon generates a protein of 526/528 amino acids truncated in domain X.

The LysM protein module is found among both prokaryotes and eukaryotes³³, and was first identified in bacterial lysin and muramidase, enzymes that degrade cell wall peptidoglycans^{30,33}. In bacteria, LysM domains are often present in proteins associated with bacterial cell walls or in proteins involved in pathogenesis. Receptor-like kinases with the LysM motif have, so far, been found only in plants but none of these receptors have been assigned a function. In Fig. 4e, the NFR1 LysM motif is aligned with the motifs from the *Volvox* chitinase, the closest *Arabidopsis* and rice receptor kinases, and the SMART consensus LysM motif. *Arabidopsis* has five receptor-like kinases with LysM domains in the predicted extracellular regions; two of these are encoded by genes with 10–11 introns, similar to the *NFR1* gene, and one of them (At3g21630)

shares 54% identity with NFR1 at the protein level. Rice has two genes in the same class, one of which encodes a protein (GenBank accession number BAB89226) that shares 32% identity with NFR1. Overall the NFR1 protein shares 33% identity with NFR5 (Supplementary Information). The two NFR1 LysM domains share conserved positions with the NFR5 LysM domains (Supplementary Information).

Expression of *NFR1*, *NFR5* and *SYMRK* genes

Because *NFR1*-dependent root hair curling in the susceptible zone located just behind the root tip predicts a localized function for the putative NFR1 receptor, we determined expression of the *NFR1* gene in different plant organs by northern blotting and real-time polymerase chain reaction (PCR) analysis. Results of both analyses demonstrate that *NFR1* gene activity is organ regulated and essentially root specific (Fig. 5a, b). Expression is not affected by inoculation with *M. loti* (Fig. 5b). In nodules, approximately 240-fold lower NFR1 transcript levels were found compared with uninoculated roots (Fig. 5a, b). Also, real-time PCR experiments revealed no difference between the levels of the two alternatively spliced *NFR1* transcripts detected in the root RNA, suggesting that the alternative splicing of intron IV is not differentially regulated (data not shown). Similar to the *NFR5* (ref. 21) and *SYMRK* (ref. 7) genes, *NFR1* is thus expressed in uninoculated roots.

Because the *NFR1*, *NFR5* and *SYMRK* receptor kinases are involved in early recognition mechanisms and *NFR1* and *NFR5* function upstream of *SYMRK*, we asked whether their transcription is interdependent. To answer this question transcript levels were determined in wild type and *nfr1-1*, *nfr5-1* and *symRK-1* mutant genotypes. The minor, approximately twofold variation of *NFR1* and *NFR5* expression in uninoculated and inoculated wild-type, *nfr1-1* and *nfr5-1* roots (Fig. 5c, d) suggests that transcriptional regulation of the two putative LysM receptor genes is mutually independent. A 20- and 10-fold lower *NFR1* steady-state transcript level in *nfr1-1* mutants 2 h and 5 h after inoculation may indicate autoregulation of expression but such a regulatory mechanism would have to be confirmed by other approaches. Comparable expression of *SYMRK* in *nfr1-1* and *nfr5-1*, as well as *NFR1* and *NFR5* in *symRK-1* (Fig. 5c–e), show that transcription of the three receptor genes in roots was independent under the conditions tested here, and that transcripts of all three genes were present in root

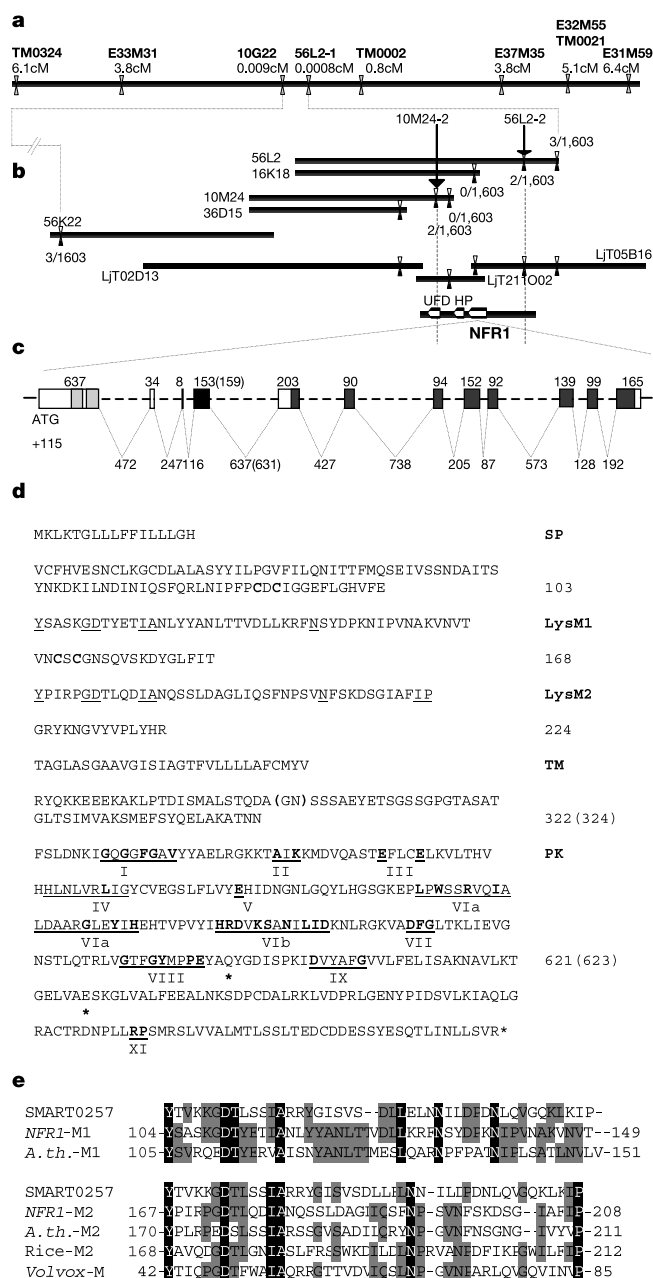


Figure 4 Positional cloning of the *NFR1* gene and domain structure of the NFR1 protein. **a**, Genetic map of the region surrounding the *NFR1* locus. Positions of the closest AFLP, microsatellite and PCR markers are given together with distances (cM). **b**, Physical map of the *NFR1* locus. BAC clones 56L2, 16K18, 10M24, 36D15, 56K22 and TAC clones LJT05B16, LJT02D13 and LJT211002 assembled to cover the region are shown. The interrupted connection signifies that the physical contig of TAC/BAC clones shown does not cover the entire region between 10G22 and 56L2-1. The numbers of recombination events detected with BAC and TAC end markers or internal markers are given. Arrows indicate positions of the two markers (10M24-2, 56L2-2) delimiting the *NFR1* locus. UFD and HP indicate the UFD1-like protein and the hypothetical protein encoded in the region. **c**, Exon-intron structure of the *NFR1* gene. Boxes, exons; LysM motifs, light grey; trans-membrane region, black; kinase domains, dark grey; dotted lines, introns; unbroken lines, 5' and 3' UTRs. The nucleotide length of all exons and introns (bp) are indicated. The numbers in parentheses indicate the alternative splicing at exon IV. **d**, Primary structure of the conceptual NFR1 protein. SP, signal peptide; LysM1/LysM2, LysM motifs; TM, transmembrane region; PK, protein kinase domains with conserved amino acids underlined and bold. Cysteine couples CxC are in bold and the LysM amino acids important for the secondary structure maintenance are underlined. The two extra amino acids resulting from alternative splicing are in parentheses. I–XI indicate kinase domains. Asterisks indicate positions of the nonsense mutations found in *nfr1-1* and *nfr1-2* mutant alleles. **e**, Alignments of the two *NFR1* LysM motifs (M1, M2) to the consensus sequences predicted by the SMART program and the *Arabidopsis* (NP_566689), rice (BAB89226) and *Volvox* (T08150) LysM motifs.

tissue up to 12 days after inoculation. This indicates that the negative autoregulation of root nodule organogenesis probably operates downstream of the receptors.

NFR1 and root cell susceptibility

Genetic complementation of the *nfr1-1* mutation in plants where only the 'hairy root' root system is transgenic demonstrated that *NFR1* functions in the root and mediates the early plant response towards rhizobial infection. On this basis we investigated whether *NFR1* has a role in the molecular processes regulating root cell susceptibility in the invasion zone. The susceptibility of root cells in this confined root segment, best characterized as the zone between emerging and fully developed root hairs, will in wild-type plants limit the number of bacterial invasions that are initiated. In response to inoculation with *M. loti*, *Lotus* *nin* mutants undergo excessive root hair deformation, with approximately a fourfold expansion of the invasion zone²⁷, suggesting a role for *NIN* as a negative regulator limiting root cell susceptibility or competence of the narrow invasion zone. In contrast, lack of root hair response to rhizobial infection or Nod-factor in *nfr1* and *nfr5* mutants predicts the positive involvement of *NFR1* and *NFR5* in these processes. We tested this hypothesis genetically in *nfr1-1 nin* double mutants (Fig. 1g, h) and in *nfr1* mutants carrying a *NIN-GUS* reporter gene whose expression marks root susceptibility. Both inoculation with *M. loti* and Nod-factor application could activate *NIN-GUS* in root hairs and epidermal cells of the invasion zone in wild-type roots (Fig. 6a) but neither treatment could induce the activity of the reporter gene in *nfr1-1* mutant roots (Fig. 6a). In a control experiment, wild-type roots were inoculated with the *M. loti* *nodC* mutant, which is incapable of synthesizing Nod-factor. Lack of detectable *GUS* reporter gene activity demonstrates that *NIN-GUS* expression is dependent on Nod-factor. As mentioned above the requirement for an upstream *NFR1* was supported by the absence of root hair deformation and curling in *nfr1-1 nin* double mutants (Fig. 1g, h). *LjCBP1-GUS*³⁴ is also rapidly activated after *M. loti* inoculation and provides a sensitive reporter of early nodulin gene expression. A parallel experiment with *LjCBP1-GUS* gave the same results as *NIN-GUS* (Fig. 6b), thus providing independent evidence supporting the requirement for functional *NFR1* in activation of the early root responses to bacterial inoculation, Nod-factor application and establishment of a functional invasion zone.

Discussion

Here we describe two transmembrane receptor kinases necessary for legume recognition of compatible rhizobial microsymbionts. The *NFR1* and *NFR5* kinase genes mediate perception of the bacterial lipochitin-oligosaccharide Nod-factor signal molecule(s), and are required for the earliest cellular and physiological responses described in legume root hairs treated with Nod-factor. Gene expression analysis suggests that epidermal root (hair) cells in the susceptible zone behind the root tip perceive the bacterial signals through *NFR1* and *NFR5* proteins, and that both establishment of a functional invasion zone and initiation of organogenesis depend on *NFR1* and *NFR5*. Continued synthesis of Nod-factors by rhizobia contained within infection threads inside the root tissue³⁵ as well as the presence of Nod-factor in root nodule cells³⁶ have previously

been shown. This implies a role for *NFR1* and *NFR5* not only during rhizobial recognition at the root epidermis/cortex but also during root tissue invasion and nodule development. The difference in expression of *NFR5* in indeterminate pea nodules versus determinate *Lotus* nodules²¹ may reflect such a requirement for *NFR5* and *NFR1* in the invasion zone or meristematic zone maintained throughout development of indeterminate nodules. More sophisticated experiments, for example using temperature-sensitive alleles, are necessary to address this question in detail.

Our physiological analysis, the gene expression studies and double mutant analysis show that *NFR1* and *NFR5* operate upstream of *SYMRK*, and that *SYMRK* operates upstream of *NIN*. A direct interaction between *NFR1* or *NFR5* and the *SYMRK* leucine-rich receptor kinase to confer rhizobial specificity on the mycorrhizal and rhizobial *SYMRK* receptor is therefore less likely.

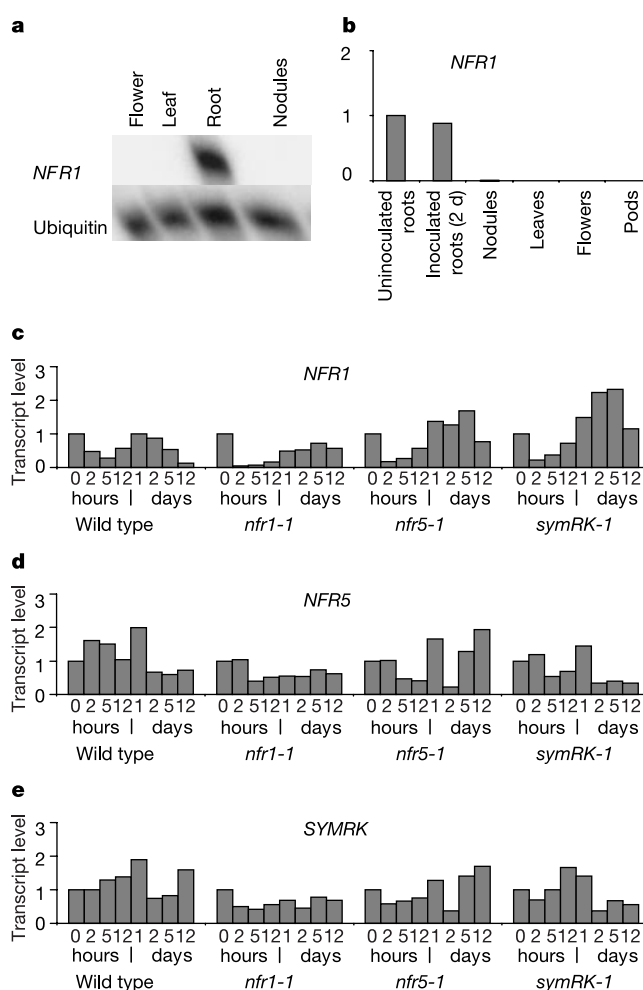


Figure 5 *NFR1*, *NFR5* and *SYMRK* gene expression. **a**, Northern hybridization visualizing elevated *NFR1* expression in root compared to flower, leaf and nodule. **b**, Steady-state levels of *NFR1* transcripts in uninoculated roots, inoculated roots, nodules, leaves, flowers and pods of wild-type plants. **c**, Transcript levels of *NFR1* in wild type, *nfr1-1*, *nfr5-1* and *symRK-1* mutant backgrounds after inoculation with *M. loti*. **d**, Transcript levels of *NFR5* in wild type, *nfr1-1*, *nfr5-1* and *symRK-1* mutant backgrounds after inoculation with *M. loti*. **e**, Transcript levels of *SYMRK* in wild type, *nfr1-1*, *nfr5-1* and *symRK-1* mutant backgrounds after inoculation with *M. loti*. Results are shown as fold increase compared to uninoculated roots at time zero. Transcript levels were measured by real-time PCR and the identity of the amplified sequences was confirmed by sequencing. ATPase was used as an internal control, and relative values normalized to the untreated roots (0 h) are shown.

Table 1 Complementation of *nfr1-1* and *nfr1-2* by the wild-type *NFR1* gene

Plant genotype	Construct	Fraction of nodulated transgenic roots	Nodules per plant
WT	Vector	14 of 15	8
<i>nfr1-1</i>	<i>NFR1</i>	62 of 103	5
<i>nfr1-1</i>	Vector	0 of 30	0
<i>nfr1-2</i>	<i>NFR1</i>	13 of 20	7.5
<i>nfr1-2</i>	Vector	0 of 7	0

The dissimilar phenotypes of *symRK* and *nfr1* and *nfr5* mutants are also in accordance with assembly of separate functional complexes, as inactive components of the same Nod-factor receptor complex would be expected to cause the same phenotype.

In our current working model, outlined in Fig. 7, NFR1 and NFR5 initiate the earliest response. Further activation or amplification of signal transduction requires the activity of SYMRK in a subsequent step shared with mycorrhizal fungi. Considering that SYMRK is a receptor-like kinase, a second signal perception step is probably involved in the activation, although an intracellular activation mechanism triggered by NFR1 and NFR5 cannot be excluded. At present, membrane depolarization, early ion fluxes and root tip swelling cannot be directly associated with individual steps in the processes linking NFR5/NFR1 and SYMRK, and these events may be in a parallel pathway. Independent expression of *NFR1*, *NFR5* and *SYMRK* is in accordance with the model. Altogether this suggests a double-lock system for rhizobial invasion of legume roots. Given the phenotypes of *nfr1* and *nfr5* mutants and the predicted NFR1 and NFR5 protein functions and topologies, we favour a simple model where NFR1 and NFR5 are assembled into a heteromeric Nod-factor receptor with their extracellular LysM domains involved in Nod-factor binding. Subsequent activation of the kinase domains would then initiate the signal transduction and gene activation cascade, ultimately leading to development of nitrogen-fixing root nodules. The different physiological response between *nfr1* and *nfr5* mutants and the apparent absence of an activation loop in NFR5 suggest that NFR5 kinase may be active immediately upon Nod-factor binding/perception. Subsequent activation (phosphorylation) of NFR1 would then initiate mobilization and activation of intracellular partners transducing the signal. Although our analysis is allele limited, the lack of physiological response in *nfr5* mutants, the attenuated alkalization in *nfr1-1* mutants and acidification in *nfr1-2* mutants support the model. Absence of NFR1 activation in *nfr5* mutants would inactivate the receptor complex, whereas the different effect of *nfr1* mutant alleles may be explained by a differential ability of the NFR1-1 and NFR1-2 allele-specific proteins to interact with intra-

cellular signal transduction components. Biochemical studies are clearly necessary to resolve the composition of the Nod-factor receptor complex(es) and the interactions with downstream signal transduction pathways.

The two transmembrane receptor kinases NFR1 and NFR5 have two (NFR1) and three (NFR5) LysM subdomains, respectively, in their extracellular domains. NMR analysis of the *Escherichia coli* MltD LysM domain³⁷ reveals a shallow groove on the surface of a $\beta\alpha\alpha\beta$ globular structure, which has been suggested as a potential peptidoglycan-binding site. Although this binding-site structure has not been confirmed experimentally, *in vivo* and *in vitro* studies of *Lactococcus lactis* autolysin demonstrate that the three LysM domains of this protein bind peptidoglycan³⁸. As both A- and B-type peptidoglycans, differing in amino acid composition as well as cross-linking, were bound, it has been concluded that autolysin LysM domains bind the *N*-acetyl-glucosamine-*N*-acetyl-mureine backbone polymer³⁸. LysM domains are frequently found together with amidase, protease or chitinase motifs³³, and two chitinases have been confirmed to carry LysM domains. One is the sex pheromone and wound-induced polypeptide from the alga *Volvox carter*, which binds and degrades chitin *in vitro*³⁹. The other is α -toxin from *Kluyveromyces lactis*, which docks on a yeast cell wall chitin receptor⁴⁰. A structure-based alignment of representative LysM domain sequences showed a pronounced variability among their primary sequence, with the exception of the amino acids directly involved in maintaining the secondary structure determined in MltD. On this background it is conceivable that the NFR1 and NFR5 LysM domains might bind Nod-factor, but a structure prediction for a putative Nod-factor-binding site needs further experimental analysis.

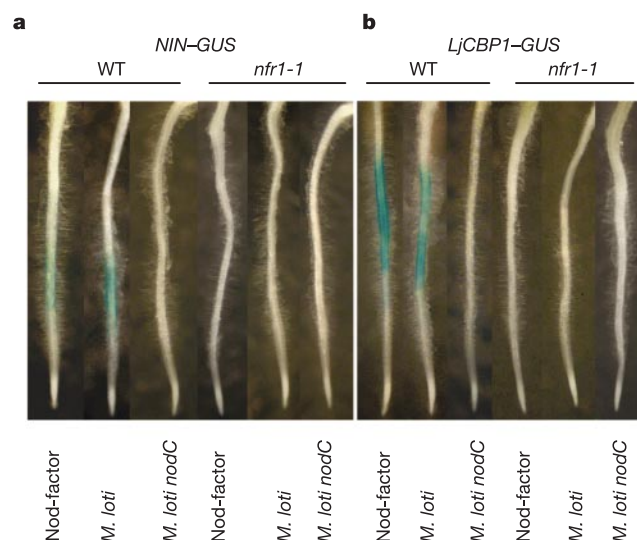


Figure 6 Induction of *NIN* gene and *LjCBP1* gene-promoter activity. **a**, Wild type and *nfr1-1* mutant roots stained for GUS activity expressed by the *NIN* gene promoter. From left to right: root treated with purified Nod-factor, root inoculated with *M. loti* and root inoculated with the *M. loti nodC* mutant impaired in Nod-factor biosynthesis. **b**, Wild type and *nfr1-1* mutant roots stained for GUS activity expressed from the *LjCBP1*-GUS reporter. From left to right: root treated with purified Nod-factor, root inoculated with *M. loti* and root inoculated with *M. loti nodC*.

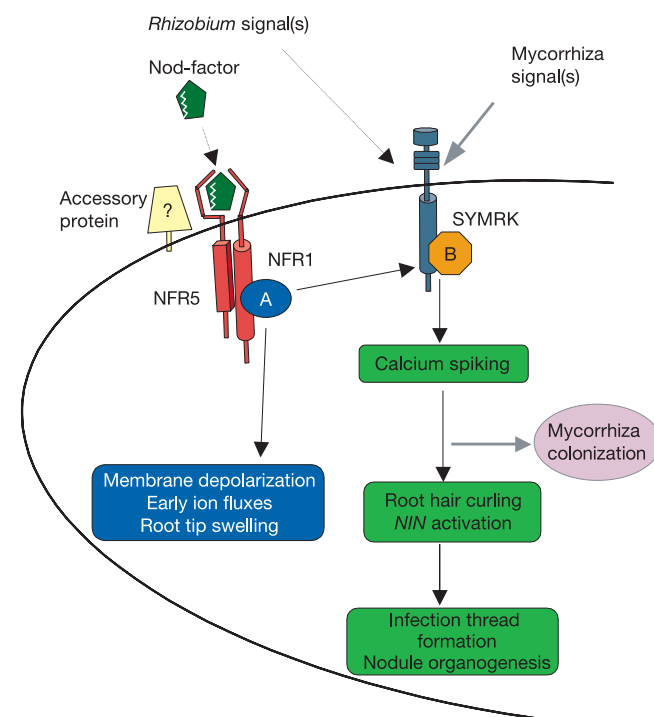


Figure 7 Working model for the functional role of NFR1 and NFR5 in plant perception of rhizobial and mycorrhizal signals. The accessory protein could correspond to the NFSB2 high-affinity Nod-factor-binding site⁴¹. A and B illustrate interacting proteins mediating downstream signalling. At present membrane depolarization, early ion fluxes and root tip swelling cannot be directly associated with individual steps in the processes linking the NFR5-NFR1 complex and SYMRK. To indicate that these events may be in a parallel pathway they are placed below the NFR1-NFR5 complex.

Physiological characterization of *nfr1* and *nfr5* mutants documents *NFR1*- and *NFR5*-mediated Nod-factor induction of early ion fluxes. The pea *sym10* mutants⁴², mutated in the orthologue of the *Lotus NFR5* gene, provide independent evidence for *NFR5* receptor kinase induction of root hair ion fluxes. Pea *sym10* mutants lack calcium spiking⁴¹, further indicating a role for *NFR5* in the later Ca^{2+} spiking response. Such a phenotype was recently described in the *Medicago truncatula nfp* mutant line⁴³. The precise role of these ion fluxes and their relation to the signal transduction process still have to be established, and our single and double mutants provide new possibilities for relating physiological events to chitin and Nod-factor perception. Application of the dual-dye-imaging spatial resolution method²³ for determining initial calcium flux and spiking responses would further substantiate the genetic requirements and timing among different legumes. Previous work demonstrating concentration-dependent Nod-factor responses in the same cell and Nod-factor structure-dependent differences in the physiological response have led to speculation about the nature and number of Nod-factor receptor(s) (ref. 23 and references therein). In this context, the inability of undecorated Nod-factor to induce the early calcium flux in *M. truncatula*²³ and membrane depolarization in alfalfa²⁴ tempts us to speculate that the specific Nod-factor response is mediated by the *NFR1*/*NFR5* receptor, whereas the *SYMRK* receptor has fewer structural requirements.

Normal mycorrhization of the *nfr1* and *nfr5* mutants demonstrates that the *NFR1* and *NFR5* receptors are not required for recognition of fungal endosymbionts but serve to link rhizobial interaction onto the common symbiotic pathway. Evolution of receptor genes may be key elements of the genetic constitution enabling legumes to establish nitrogen-fixing symbiosis. Transfer of these receptor genes to plants only capable of mycorrhizal symbiosis, and introduction of rhizobia-induced signal transduction through the common pathway, would indicate how far the nodule developmental process would progress in non-nodulating plants such as cereals. □

Methods

Plant material and mapping population

Isolation of the *nfr1-1* and *nfr1-2* (previously called *sym1-1* and *sym1-2*) mutants has been described previously¹⁹. Plants were grown with or without *M. loti* NZP2235 as previously outlined⁴⁴. An F_2 mapping population was established by crossing a *L. japonicus* 'Gifu' *nfr1-1* mutant to wild-type *L. japonicus* ecotype 'MG-20'. F_2 plants homozygous for the *nfr1-1* mutant allele were identified after screening for the non-nodulation mutant phenotype. In total 1,603 homozygous F_2 mutant plants were analysed.

Positional cloning and complementation

Using bulked segregant analysis and AFLP the *NFR1* locus was initially mapped to a 7.6-cM region between AFLP markers E33M31 and E37M35 on chromosome I. Subsequently, microsatellite markers (TM markers) and single nucleotide polymorphic markers developed from BAC or TAC clones anchored to the general genetic map of the region were used for fine mapping and for building the physical TAC/BAC contig. A contig of approximately 400 kb was assembled to cover the *NFR1* locus between the two flanking markers 56L2-1 and 10G22. Additional fine mapping using TAC and BAC sequence physically delimited the *NFR1* locus to 250 kb between the marginal recombination breakpoints 56L2-1 and 56K22. On the basis of sequence differences between the two parents, additional PCR markers were developed inside this contig and the region was narrowed down to 36 kb between markers 56L2-2 and 10M24-2. Finally, LjT05B16, LjT02D13 and LjT211O02 containing the entire *NFR1* region were sequenced as part of the *Lotus* genome sequencing programme⁴⁵ together with the *NFR1* gene region from 56L2.

For *in planta* complementation a 9.2-kb genomic fragment containing the *NFR1* gene, 3,016-bp promoter region and 392-bp 3' UTR was used. The fragment was subcloned from BAC 56L2 and inserted into pV10 plasmid. A tri-parental mating was performed to integrate the construct into a *rhizogenes* strain AR12 and AR1193. Transgenic hairy roots and nodulation tests were done as previously described²⁹.

cDNA isolation

A partial cDNA clone of 1,068 bp corresponding to the 3' end of *NFR1* was isolated by screening a λ ZAPII cDNA library prepared from mRNA extracted from inoculated roots. The missing 5' end was amplified by 5' RACE PCR on RNA from uninoculated root following the manufacturers instruction (Clontech). The 2.2-kb 5' RACE product was cloned into the pCR2.1 vector and nine clones were sequenced to identify the transcription start point. The full cDNA sequence was assembled from the cDNA and the 5' RACE product.

Expression analysis

Total RNA was isolated from root, leaf, flower, pod and nodule tissues of uninoculated or inoculated plants and purified through a CsCl cushion. Twenty micrograms total RNA was used for northern analysis with an *NFR1*-specific, 1,280-bp probe covering the extracellular domain (position 4045–5324 amplified by 5'-TAATTATCAGAGTAAGTGTGAC-3', 5'-AGTTACCCACCTGTGGTAC-3'). Transcript levels after inoculation were determined by quantitative real-time RT-PCR on poly(A)⁺ mRNA purified using the Dynabeads mRNA direct kit. All RNA samples were tested for contaminating DNA using PCR primers specific for the untranscribed *NIN* gene promoter (5'-GTTTTCAAGAATGGGAGGGG-3', 5'-CTCCTCTGGTTTCATTGGTG-3'). If necessary, RNA was incubated with DNaseI, treated with phenol:chloroform and retested for contaminating DNA. First strand cDNA was prepared using reverse transcriptase (Roche) and RACE dT primer. Quantitative PCR was performed on a LightCycler (Roche Molecular Biochemicals) using FastStart DNA master SYBR greenI kit (Roche) to amplify the target transcripts from 5 μ l diluted cDNA. The same cDNA pool was used for amplification of all tested transcripts. The relative quantification software (Roche) was used to determine the efficiency-corrected relative transcript concentration, normalized to a calibrator sample. ATP synthase was used as a reference gene⁴⁶ (AW19841). ATP synthase has the same constitutive expression profile as ubiquitin but is expressed at a level that is approximately 25-fold lower (our own unpublished data). Each RNA sample was again tested for DNA using the *NIN* gene primers in real-time PCR reactions. Melting curve analysis and sequencing of the amplified products was used to determine their identity. Intron-spanning primers were used for ATP synthase, *NFR1* and *NIN*. The primers used for transcript amplification were: ATP synthase, 5'-CAATGTCGCGCAAGGCCCATGGTG-3', 5'-AACACCACTCTCGATCATTTCTCTG-3'; *NFR1*, 5'-CCCTTGATACACAGAAC-3', 5'-GCTTCTCTCTCTCTCTCTCTCTG-3'; *NFR5*, 5'-GCAAGGGAAGTAATTCAG-3', 5'-CAGACAACCTCTCCATC-3'; *SYMRK*, 5'-TGAATGGAATCTTTTGATTGCG-3', 5'-TCACCTGACCTTGACCCAGA-3'; *NIN*, 5'-AATGCTCTTGATCAGGCTG-3', 5'-AGGAGCCCAAGTGAGTGCTA-3'; *ENOD2*, 5'-CAGAAAAACCACCCTGT-3', 5'-ATGAGGCGAATACACTGGTG-3'. Representative data from one of three equivalent experiments are shown.

Root hair curling and GUS assays

Lotus Gifu wild type, *nfr1-1*, *nfr1-2*, *symRK-3* and *nin-1* mutants, and *nfr1-2 symRK-3* and *nfr1-2 nin-1* double mutant seeds were sterilized and germinated for 3 days. Seedlings were transferred on one-quarter B&D medium for the next 3 days, and then inoculated with 20 μ l of a 1:100 dilution of a 2-day-old *M. loti* NZP2235 culture (optical density at 600 nm of 0.85) with sand coated with Nod-factor from *M. loti* strain R7A, or with sterile water as a control. Root hairs were microscopically analysed for specific deformations in a minimum of 30 seedlings. *NIN-GUS* and *LjCBP1-GUS* activity in wild type, *nfr1-1* and *nfr5-1* mutants was assayed according to ref. 47. Nod-factor was extracted from bacterial cultures of *M. loti* strain R7A and purified by high-pressure liquid chromatography as previously described¹³. The structure of these Nod-factors is identical to those purified and characterized from strain EIR¹³.

Electrophysiology

pH was monitored continuously in a flow-through regime using a pH-selective microelectrode. The microelectrode was placed within the root hair space of the zone that is competent for rhizobial invasion and responds to *Rhizobium* and Nod-factor, as previously described^{18,22,24}. Wherever shown, membrane potential was measured simultaneously with pH. For each genotype the number of pH and E_m measurements are given as (pH, E_m): WT (8, 6); *nfr5-1* (5, 3); *nfr5-2* (3, 3); *nfr1-1* (3, 2); *nfr1-2* (3, 2); *symRK-3* (3, 3); *symRK-1* (3, 3); *nfr1-2 symRK-3* (3, 4); CH-O, 3 or more for both measurements.

Computer analysis

Sequences were analysed by BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) and PredictProtein (<http://www.embl-heidelberg.de/predictprotein/predictprotein.html>). The signal peptide was predicted by SignalP v.1.1 (<http://www.cbs.dtu.dk/services/SignalP-2.0/>) and the transmembrane region was predicted by TMHMM v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>).

Received 16 July; accepted 11 September 2003; doi:10.1038/nature02039.

1. Parniske, M. Intracellular accommodation of microbes by plants: a common developmental program for symbiosis and disease? *Curr. Opin. Plant Biol.* **3**, 320–328 (2000).
2. Gadkar, V., David-Schwartz, R., Kunik, T. & Kapulnik, Y. Arbuscular mycorrhizal fungal colonization. Factors involved in host recognition. *Plant Physiol.* **127**, 1493–1499 (2001).
3. Szczygłowski, K. & Amyot, L. Symbiosis, inventiveness by recruitment? *Plant Physiol.* **131**, 935–940 (2003).
4. Duc, G., Trouvelot, A., Gianinazzi-Pearson, V. & Gianinazzi, S. First report of non-mycorrhizal plant mutants (Myc⁻) obtained in pea (*Pisum sativum* L.) and fababean (*Vicia faba* L.). *Plant Sci.* **60**, 215–222 (1989).
5. Wegel, E., Schauser, L., Sandal, N., Stougaard, J. & Parniske, M. Mycorrhiza mutants of *Lotus japonicus* define genetically independent steps during symbiotic infection. *Mol. Plant Microbe Interact.* **11**, 933–936 (1998).
6. Stougaard, J. Genetics and genomics of root symbiosis. *Curr. Opin. Plant Biol.* **4**, 328–335 (2001).
7. Stracke, S. *et al.* A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* **417**, 959–962 (2002).
8. Endre, G. *et al.* A receptor kinase gene regulating symbiotic nodule development. *Nature* **417**, 962–966 (2002).
9. Kosuta, S. *et al.* A diffusible factor from arbuscular mycorrhizal fungi induces symbiosis-specific *MtENOD11* expression in roots of *Medicago truncatula*. *Plant Physiol.* **131**, 952–962 (2003).
10. Lerouge, P. *et al.* Symbiotic host specificity of *Rhizobium meliloti* is determined by a sulphated and

- acetylated glucosamine oligosaccharide signal. *Nature* **344**, 781–784 (1990).
11. Spaink, H. P. *et al.* A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. *Nature* **354**, 125–130 (1991).
12. Truchet, G. *et al.* Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. *Nature* **351**, 670–673 (1991).
13. Lopez-Lara, I. M. *et al.* Structural identification of the lipo-chitin oligosaccharide nodulation signals of *Rhizobium loti*. *Mol. Microbiol.* **15**, 627–638 (1995).
14. Niwa, S. *et al.* Responses of a model legume *Lotus japonicus* to lipochitin oligosaccharide nodulation factors purified from *Mesorhizobium loti* JRL501. *Mol. Plant Microbe Interact.* **14**, 848–856 (2001).
15. Kistner, C. & Parniske, M. Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* **7**, 511–518 (2002).
16. Pacios-Bras, C. *et al.* A *Lotus japonicus* nodulation system based on heterologous expression of the fucosyl transferase *NodZ* and the acetyl transferase *Noll* in *Rhizobium leguminosarum*. *Mol. Plant Microbe Interact.* **13**, 475–479 (2000).
17. Ehrhardt, D. W., Atkinson, E. M. & Long, S. R. Depolarization of alfalfa root hair membrane potential by *Rhizobium meliloti* Nod factors. *Science* **256**, 998–1000 (1992).
18. Felle, H. H., Kondorosi, E., Kondorosi, A. & Schultze, M. The role of ion fluxes in Nod factor signalling in *Medicago sativa*. *Plant J.* **13**, 455–463 (1998).
19. Schauser, L. *et al.* Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. *Mol. Gen. Genet.* **259**, 414–423 (1998).
20. Szczygłowski, K. *et al.* Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Mol. Plant Microbe Interact.* **11**, 684–697 (1998).
21. Madsen, E. B. *et al.* A receptor-kinase gene of the LysM type is involved in perception of rhizobial signals. *Nature* **425**, 637–640 (2003).
22. Felle, H. H., Kondorosi, E., Kondorosi, A. & Schultze, M. Nod signal-induced plasma membrane potential changes in alfalfa root hairs are differentially sensitive to structural modifications of the lipochitoligosaccharide. *Plant J.* **7**, 939–947 (1995).
23. Shaw, S. L. & Long, S. R. Nod factor elicits two separable calcium responses in *Medicago truncatula* root hair cells. *Plant Physiol.* **131**, 976–984 (2003).
24. Felle, H. H., Kondorosi, E., Kondorosi, A. & Schultze, M. How alfalfa root hairs discriminate between Nod factors and oligochitin elicitors. *Plant Physiol.* **124**, 1373–1380 (2000).
25. Govers, F. *et al.* Characterization of the pea ENOD12B gene and expression analyses of the two ENOD12 genes in nodule, stem and flower tissue. *Mol. Gen. Genet.* **228**, 160–166 (1991).
26. Colebatch, G. *et al.* Novel aspects of symbiotic nitrogen fixation uncovered by transcript profiling with cDNA arrays. *Mol. Plant Microbe Interact.* **15**, 411–420 (2002).
27. Schauser, L., Roussis, A., Stiller, J. & Stougaard, J. A plant regulator controlling development of symbiotic root nodules. *Nature* **402**, 191–195 (1999).
28. Sandal, N. *et al.* A genetic linkage map of the model legume *Lotus japonicus* and strategies for fast mapping of new loci. *Genetics* **161**, 1673–1683 (2002).
29. Stougaard, J. *Agrobacterium rhizogenes* as a vector for transforming higher plants. *Methods Mol. Biol.* **49**, 49–61 (1995).
30. Joris, B. *et al.* Modular design of the *Enterococcus hirae* muramidase-2 and *Streptococcus faecalis* autolysin. *FEMS Microbiol. Lett.* **70**, 257–264 (1992).
31. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).
32. Schenk, P. W. & Snaar-Jagalska, B. E. Signal perception and transduction: the role of protein kinases. *Biochim. Biophys. Acta* **1449**, 1–24 (1999).
33. Pontig, C. P. *et al.* Eukaryotic signalling domain homologous in Archaea and Bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**, 729–745 (1999).
34. Webb, J. *et al.* *Mesorhizobium loti* increases root-specific expression of a calcium-binding protein homologue identified by promoter tagging in *Lotus japonicus*. *Mol. Plant Microbe Interact.* **13**, 606–616 (2000).
35. Schlaman, H. R. *et al.* Suppression of nodulation gene expression in bacteroids of *Rhizobium leguminosarum* biovar *viceae*. *J. Bacteriol.* **173**, 4277–4287 (1991).
36. Timmers, A. C. J., Auriac, M.-C., de Billy, F. & Truchet, G. Nod factor internalization and microtubular cytoskeleton changes occur concomitantly during nodule differentiation in alfalfa. *Development* **125**, 339–349 (1998).
37. Bateman, A. & Bycroft, M. The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J. Mol. Biol.* **299**, 1113–1119 (2000).
38. Steen, A. *et al.* Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *J. Biol. Chem.* **278**, 23874–23881 (2003).
39. Amon, P., Haas, E. & Sumper, M. The sex-inducing pheromone and wounding trigger the same set of genes in the multicellular green alga *Volvox*. *Plant Cell* **10**, 781–789 (1998).
40. Butler, A. R., O'Donnell, R. W., Martin, V. J., Gooday, G. W. & Stark, M. J. *Kluyveromyces lactis* toxin has an essential chitinase activity. *Eur. J. Biochem.* **199**, 483–488 (1991).
41. Gressent, F. *et al.* Ligand specificity of a high-affinity binding site for lipo-chitooligosaccharidic Nod factors in *Medicago* cell suspension cultures. *Proc. Natl Acad. Sci. USA* **96**, 4704–4709 (1999).
42. Walker, S. A., Viprey, V. & Downie, J. A. Dissection of nodulation signaling using pea mutants defective for calcium spiking induced by Nod factors and chitin oligomers. *Proc. Natl Acad. Sci. USA* **97**, 13413–13418 (2000).
43. Amor, B. B. *et al.* The NFP locus of *Medicago truncatula* controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *Plant J.* **34**, 495–506 (2003).
44. Handberg, K. & Stougaard, J. *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J.* **2**, 487–496 (1992).
45. Asamizu, E. *et al.* Structural analysis of a *Lotus japonicus* genome. IV. Sequence features and mapping of seventy-three TAC clones which cover the 7.5-Mb regions of the genome. *DNA Res.* **10**, 115–122 (2003).
46. Gerard, C. J., Andrejka, L. M. & Macina, R. A. Mitochondrial ATP synthase 6 as an endogenous control in the quantitative RT-PCR analysis of clinical cancer samples. *Mol. Diagn.* **5**, 39–46 (2000).
47. Kosugi, S., Ohashi, Y., Nakajima, K. & Arai, Y. An improved assay for β -glucuronidase in transformed cells: Methanol almost completely suppresses a putative endogenous β -glucuronidase activity. *Plant Sci.* **70**, 133–140 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Webb for making the *LjCBP1*–*GUS* line available. S.R. was supported by an EU Marie Curie Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S. (stougaard@mb.au.dk). The sequences for the *L. japonicus* ecotype Gifu *NFR1* gene and mRNA are deposited in the EMBL database under accession numbers AJ575246, AJ575247, AJ575248 and AJ575249. The MG20 ecotype TAC sequences are deposited in GenBank under accession numbers LJT02D13 AP006535, LJT05B16 AP006536 and LJT211O02 AP006537.

Dodecahedral space topology as an explanation for weak wide-angle temperature correlations in the cosmic microwave background

Jean-Pierre Luminet¹, Jeffrey R. Weeks², Alain Riazuelo³, Roland Lehoucq^{1,3} & Jean-Philippe Uzan⁴

¹Observatoire de Paris, 92195 Meudon Cedex, France

²15 Farmer Street, Canton, New York 13617-1120, USA

³CEA/Saclay, 91191 Gif-sur-Yvette Cedex, France

⁴Laboratoire de Physique Théorique, Université Paris XI, 91405 Orsay Cedex, France

The current 'standard model' of cosmology posits an infinite flat universe forever expanding under the pressure of dark energy. First-year data from the Wilkinson Microwave Anisotropy Probe (WMAP) confirm this model to spectacular precision on all but the largest scales^{1,2}. Temperature correlations across the microwave sky match expectations on angular scales narrower than 60° but, contrary to predictions, vanish on scales wider than 60°. Several explanations have been proposed^{3,4}. One natural approach questions the underlying geometry of space—namely, its curvature⁵ and topology⁶. In an infinite flat space, waves from the Big Bang would fill the universe on all length scales. The observed lack of temperature correlations on scales beyond 60° means that the broadest waves are missing, perhaps because space itself is not big enough to support them. Here we present a simple geometrical model of a finite space—the Poincaré dodecahedral space—which accounts for WMAP's observations with no fine-tuning required. The predicted density is $\Omega_0 \approx 1.013 > 1$, and the model also predicts temperature correlations in matching circles on the sky⁷.

Temperature fluctuations on the microwave sky may be expressed as a sum of spherical harmonics, just as music and other sounds may be expressed as a sum of ordinary harmonics. A musical note is the sum of a fundamental, a second harmonic, a third harmonic, and so on. The relative strengths of the harmonics—the note's spectrum—determines the tone quality, distinguishing, say, a sustained middle C played on a flute from the same note played on a clarinet. Analogously, the temperature map on the microwave sky is the sum of spherical harmonics. The relative strengths of the harmonics—the power spectrum—is a signature of the physics and geometry of the Universe. Indeed, the power spectrum is the primary tool researchers use to test their models' predictions against observed reality.

The infinite universe model gets into trouble at the low end of the

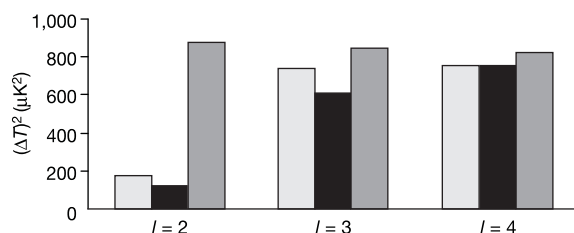


Figure 1 Comparison of the WMAP power spectrum to that of Poincaré dodecahedral space and an infinite flat universe. At the low end of the power spectrum, WMAP's results (black bars) match the Poincaré dodecahedral space (light grey) better than they match the expectations for an infinite flat universe (dark grey). Computed for $\Omega_m = 0.28$ and $\Omega_\Lambda = 0.734$ with Poincaré space data normalized to the $l = 4$ term.

power spectrum (Fig. 1). The lowest harmonic—the dipole, with wavenumber $l = 1$ —is unobservable because the Doppler effect of the Solar System's motion through space creates a dipole 100 times stronger, swamping out the underlying cosmological dipole. The first observable harmonic is the quadrupole, with wavenumber $l = 2$. WMAP found a quadrupole only about one-seventh as strong as would be expected in an infinite flat space. The probability that this could happen by mere chance has been estimated at about 0.2% (ref. 2). The octopole term, with wavenumber $l = 3$, is also weak at 72% of the expected value, but not nearly so dramatic or significant as the quadrupole. For large values of l , ranging up to $l = 900$ and corresponding to small-scale temperature fluctuations, the spectrum tracks the infinite universe predictions exceedingly well.

Cosmologists thus face the challenge of finding a model that accounts for the weak quadrupole while maintaining the success of the infinite flat universe model on small scales (high l). The weak wide-angle temperature correlations discussed in the introductory paragraph correspond directly to the weak quadrupole.

Microwave background temperature fluctuations arise primarily (but not exclusively) from density fluctuations in the early Universe, because photons travelling from denser regions do a little extra work against gravity and therefore arrive cooler, while photons from less dense regions do less work against gravity and arrive warmer. The density fluctuations across space split into a sum of three-dimensional harmonics—in effect, the vibrational overtones of space itself—just as temperature fluctuations on the sky split into a sum of two-dimensional spherical harmonics and a musical note splits into a sum of one-dimensional harmonics. The low quadrupole implies a cut-off on the wavelengths of the three-dimensional harmonics. Such a cut-off presents an awkward problem in infinite flat space, because it defines a preferred length scale in an otherwise scale-invariant space. A more natural explanation invokes a finite universe, where the size of space itself imposes a cut-off on the wavelengths (Fig. 2). Just as the vibrations of a bell cannot be larger than the bell itself, the density fluctuations in space cannot be larger than space itself. Whereas most potential spatial topologies fail to fit the WMAP results, the Poincaré dodecahedral space fits them very well.

The Poincaré dodecahedral space is a dodecahedral block of space with opposite faces abstractly glued together, so objects passing out of the dodecahedron across any face return from the opposite face. Light travels across the faces in the same way, so if we sit inside the

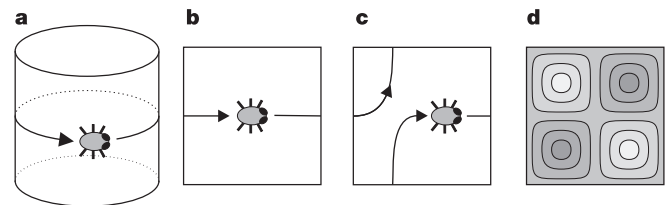


Figure 2 Wavelengths of density fluctuations are limited by the size of a finite 'wraparound' universe. **a**, A two-dimensional creature living on the surface of a cylinder travels due east, eventually going all the way around the cylinder and returning to her starting point. **b**, If we cut the cylinder open and flatten it into a square, the creature's path goes out of the square's right side and returns from the left side. **c**, A flat torus is like a cylinder, only now the top and bottom sides connect as well as the left and right. **d**, Waves in a torus universe may have wavelengths no longer than the width of the square itself. To construct a multiconnected three-dimensional space, start with a solid polyhedron (for example, a cube) and identify its faces in pairs, so that any object leaving the polyhedron through one face returns from the matching face. Such a multiconnected space supports standing waves whose exact shape depends on both the geometry of the polyhedron and how the faces are identified. Nevertheless, the same principle applies, that the wavelength cannot exceed the size of the polyhedron itself. In particular, the inhabitants of such a space will observe a cut-off in the wavelengths of density fluctuations.

dodecahedron and look outward across a face, our line of sight re-enters the dodecahedron from the opposite face. We have the illusion of looking into an adjacent copy of the dodecahedron. If we take the original dodecahedral block of space not as a euclidean dodecahedron (with edge angles $\sim 117^\circ$) but as a spherical dodecahedron (with edge angles exactly 120°), then adjacent images of the dodecahedron fit together snugly to tile the hypersphere (Fig. 3b), analogously to the way adjacent images of spherical pentagons (with perfect 120° angles) fit snugly to tile an ordinary sphere (Fig. 3a).

The power spectrum of the Poincaré dodecahedral space depends strongly on the assumed mass-energy density parameter Ω_0 (Fig. 4). The octopole term ($l=3$) matches WMAP's octopole best when $1.010 < \Omega_0 < 1.014$. Encouragingly, in the subinterval $1.012 < \Omega_0 < 1.014$ the quadrupole ($l=2$) also matches the WMAP value. More encouragingly still, this subinterval agrees well with observations, falling comfortably within WMAP's best-fit range of $\Omega_0 = 1.02 \pm 0.02$ (ref. 1).

The excellent agreement with WMAP's results is all the more striking because the Poincaré dodecahedral space offers no free

parameters in its construction. The Poincaré space is rigid, meaning that geometrical considerations require a completely regular dodecahedron. By contrast, a 3-torus, which is nominally made by gluing opposite faces of a cube but may be freely deformed to any parallelepiped, has six degrees of freedom in its geometrical construction. Furthermore, the Poincaré space is globally homogeneous, meaning that its geometry—and therefore its power spectrum—looks statistically the same to all observers within it. By contrast, a typical finite space looks different to observers sitting at different locations.

Confirmation of a positively curved universe ($\Omega_0 > 1$) would require revisions to current theories of inflation, but it is not certain how severe those changes would be. Some researchers argue that positive curvature would not disrupt the overall mechanism and effects of inflation, but only limit the factor by which space expands during the inflationary epoch to about a factor of ten⁸. Others claim that such models require fine-tuning and are less natural than the infinite flat space model⁹.

Having accounted for the weak observed quadrupole, the Poincaré dodecahedral space will face two more experimental tests in the next few years. (1) The Cornish–Spergel–Starkman circles-in-the-sky method⁷ predicts temperature correlations along matching circles in small multiconnected spaces such as this one. When $\Omega_0 \approx 1.013$ the horizon radius is about 0.38 in units of the curvature radius, while the dodecahedron's inradius and outradius are 0.31 and 0.39, respectively, in the same units. In this case the horizon sphere self-intersects in six pairs of circles of angular radius about 35° , making the dodecahedral space a good candidate for circle detection if technical problems (galactic foreground removal, integrated Sachs–Wolfe effect, Doppler effect of plasma motion) can be overcome. Indeed, the Poincaré dodecahedral space makes circle searching easier than in the general case, because the six pairs of matching circles must a priori lie in a symmetrical pattern like the faces of a dodecahedron, thus allowing the searcher to slightly relax the noise tolerances without increasing the danger of a false positive. (2) The Poincaré dodecahedral space predicts $\Omega_0 \approx 1.013 > 1$. The upcoming Planck Surveyor data (or possibly even the existing WMAP data in conjunction with other data sets) should determine Ω_0 to within 1%. Finding $\Omega_0 < 1.01$ would refute the Poincaré space as a cosmological model, while $\Omega_0 > 1.01$ would provide strong evidence in its favour.

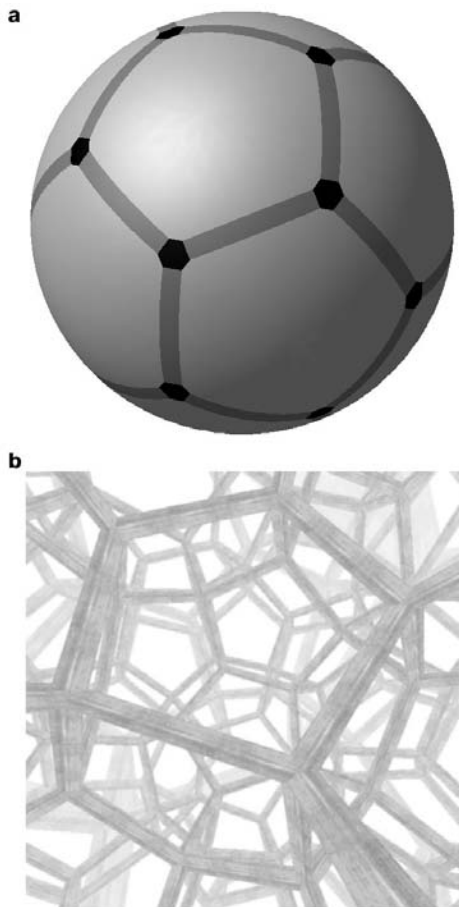


Figure 3 Spherical pentagons and dodecahedra fit snugly, unlike their euclidean counterparts. **a**, 12 spherical pentagons tile the surface of an ordinary sphere. They fit together snugly because their corner angles are exactly 120° . Note that each spherical pentagon is just a pentagonal piece of a sphere. **b**, 120 spherical dodecahedra tile the surface of a hypersphere. A hypersphere is the three-dimensional surface of a four-dimensional ball. Note that each spherical dodecahedron is just a dodecahedral piece of a hypersphere. The spherical dodecahedra fit together snugly because their edge angles are exactly 120° . In the construction of the Poincaré dodecahedral space, the dodecahedron's 30 edges come together in ten groups of three edges each, forcing the dihedral angles to be 120° and requiring a spherical dodecahedron rather than a euclidean one. Software for visualizing spherical dodecahedra and the Poincaré dodecahedral space is available at (<http://www.geometrygames.org/CurvedSpaces>).

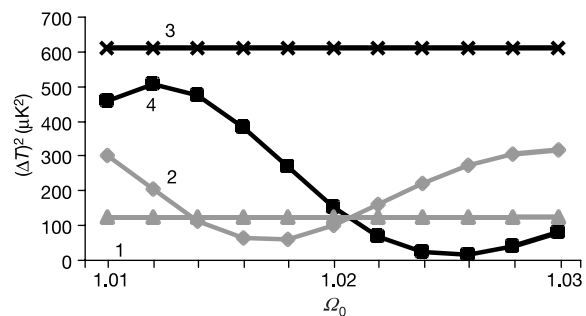


Figure 4 Values of the mass-energy density parameter Ω_0 for which the Poincaré dodecahedral space agrees with WMAP's results. The Poincaré dodecahedral space quadrupole (trace 2) and octopole (trace 4) fit the WMAP quadrupole (trace 1) and octopole (trace 3) when $1.012 < \Omega_0 < 1.014$. Larger values of Ω_0 predict an unrealistically weak octopole. To obtain these predicted values, we first computed the eigenmodes of the Poincaré dodecahedral space using the 'ghost method' of ref. 10 with two of the matrix generators computed in Appendix B of ref. 11, and then applied the method of ref. 12, using $\Omega_m = 0.28$ and $\Omega_\Lambda = \Omega_0 - 0.28$, to obtain a power spectrum and to simulate sky maps. Numerical limitations restricted our set of three-dimensional eigenmodes to wavenumbers $k < 30$, which in turn restricted the reliable portion of the power spectrum to $l = 2, 3, 4$. We set the overall normalization factor to match the WMAP data at $l = 4$ and then examined the predictions for $l = 2, 3$.

Since antiquity, humans have wondered whether our Universe is finite or infinite. Now, after more than two millennia of speculation, observational data might finally settle this ancient question. □

Received 23 June; accepted 28 July 2003; doi:10.1038/nature01944.

1. Bennett, C. L. *et al.* First year Wilkinson Microwave Anisotropy Probe (WMAP 1) observations: Preliminary maps and basic results. *Astrophys. J. Suppl.* **148**, 1–27 (2003).
2. Spergel, D. N. *et al.* First year Wilkinson Microwave Anisotropy Probe (WMAP 1) observations: Determination of cosmological parameters. *Astrophys. J. Suppl.* **148**, 175–194 (2003).
3. Contaldi, C. R., Peloso, M., Kofman, L. & Linde, A. Suppressing the lower multipoles in the CMB anisotropies. *J. Cosmol. Astropart. Phys.* **07**, 002 (2003).
4. Cline, J. M., Crotty, P. & Lesgourgues, J. Does the small CMB quadrupole moment suggest new physics? *J. Cosmol. Astropart. Phys.* (in the press); preprint at (<http://arXiv.org/astro-ph/0304558>) (2003).
5. Efstathiou, G. Is the low CMB quadrupole a signature of spatial curvature? *Mon. Not. R. Astron. Soc.* **343**, L95–L98 (2003).
6. Tegmark, M., de Oliveira-Costa, A. & Hamilton, A. A high resolution foreground cleaned CMB map from WMAP. *Phys. Rev. D* (in the press); preprint at (<http://arXiv.org/astro-ph/0302496>) (2003).
7. Cornish, N., Spergel, D. & Starkman, G. Circles in the sky: Finding topology with the microwave background radiation. *Class. Quant. Grav.* **15**, 2657–2670 (1998).
8. Uzan, J.-P., Kirchner, U. & Ellis, G. F. R. WMAP data and the curvature of space. *Mon. Not. R. Astron. Soc.* (in the press).
9. Linde, A. Can we have inflation with $\Omega > 1$? *J. Cosmol. Astropart. Phys.* **05**, 002 (2003).
10. Lehoucq, R., Weeks, J., Uzan, J.-P., Gausmann, E. & Luminet, J.-P. Eigenmodes of 3-dimensional spherical spaces and their application to cosmology. *Class. Quant. Grav.* **19**, 4683–4708 (2002).
11. Gausmann, E., Lehoucq, R., Luminet, J.-P., Uzan, J.-P. & Weeks, J. Topological lensing in spherical spaces. *Class. Quant. Grav.* **18**, 5155–5186 (2001).
12. Riazuelo, A., Uzan, J.-P., Lehoucq, R. & Weeks, J. Simulating cosmic microwave background maps in multi-connected spaces. *Phys. Rev. D* (in the press); preprint at (<http://arXiv.org/astro-ph/0212223>) (2002).

Acknowledgements J.R.W. thanks the MacArthur Foundation for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.W. (weeks@northnet.org).

Fermi-liquid breakdown in the paramagnetic phase of a pure metal

N. Doiron-Leyraud¹, I. R. Walker¹, L. Taillefer², M. J. Steiner¹, S. R. Julian¹ & G. G. Lonzarich¹

¹Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK

²Département de Physique, Université de Sherbrooke, Québec J1K 2R1, Canada

Fermi-liquid theory¹ (the standard model of metals) has been challenged by the discovery of anomalous properties in an increasingly large number of metals. The anomalies often occur near a quantum critical point—a continuous phase transition in the limit of absolute zero, typically between magnetically ordered and paramagnetic phases. Although not understood in detail, unusual behaviour in the vicinity of such quantum critical points was anticipated nearly three decades ago by theories going beyond the standard model^{2–5}. Here we report electrical resistivity measurements of the 3d metal MnSi, indicating an unexpected breakdown of the Fermi-liquid model—not in a narrow crossover region close to a quantum critical point^{6,7} where it is normally expected to fail, but over a wide region of the phase diagram near a first-order magnetic transition. In this regime, corrections to the Fermi-liquid model are expected to be small. The range in pressure, temperature and applied magnetic field over which we observe an anomalous temperature dependence of the electrical resistivity in MnSi is not consistent with the crossover behaviour widely seen in quantum critical systems^{8,9,31}. This may suggest the emergence of a well defined but enigmatic quantum phase of matter.

In the investigation of anomalous metallic properties, often referred to as non-Fermi-liquid phenomena, complications of various types may arise that can lead to ambiguous interpretations. These can be related, for example, to band structure anomalies, to magneto-crystalline anisotropies, to the effective dimensionality or to disorder effects. In other materials, the emergence of superconductivity may make it difficult to probe the fundamental nature of the normal state of the underlying electronic system. In order to conclusively assess and clarify our ideas about non-Fermi-liquid phenomena in metals, it is essential to study materials that fulfil the desirable criteria of purity, simplicity and convenience. The itinerant-electron ferromagnet MnSi appears to be one such example¹⁰. Under ambient conditions of pressure and applied magnetic field, MnSi orders ferromagnetically with a small average moment of up to 0.4 μ_B per Mn atom (where μ_B is the Bohr magneton) and a Curie temperature T_C of 29.5 K (ref. 11). At low temperatures, it exhibits properties consistent with the formation of a weakly spin-polarized Fermi liquid dominated by moderately renormalized 3d bands. With its full three-dimensional cubic crystal structure, the electronic and magnetic properties of MnSi are essentially isotropic, except for the effects of a well-understood long-wavelength helical twist of the ferromagnetic order characteristic of crystalline structures lacking inversion symmetry—such as the B20 lattice of MnSi (ref. 12).

The possibility of producing ultra-pure single crystals of MnSi with a disorder mean free path in excess of 5,000 Å (as confirmed by

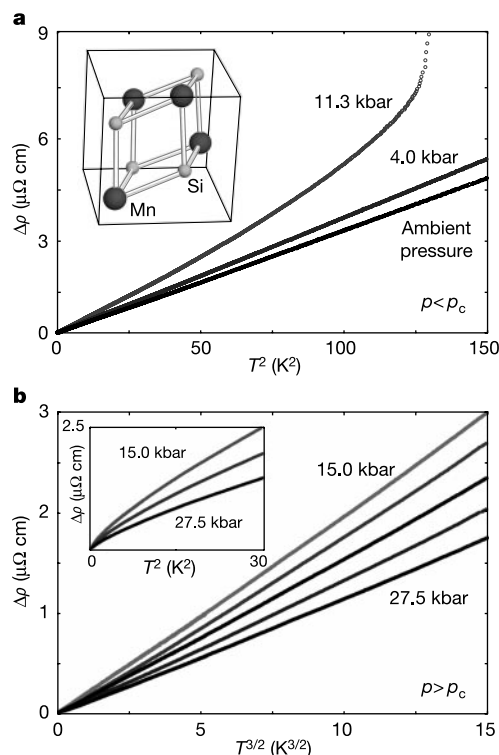


Figure 1 Dependence on temperature of the electrical resistivity of MnSi. The temperature-dependent part $\Delta\rho$ of the resistivity above the residual resistivity $\rho_0 = 0.17 \mu\Omega \text{ cm}$ is plotted for different conditions of pressure. **a**, $\Delta\rho$ on a quadratic temperature scale up to about 12 K and for pressures below the critical pressure, that is, in the ferromagnetic phase. As the Curie temperature is suppressed with pressure, the quadratic regime is observed only below about 6 K at 11.3 kbar. Inset, the cubic unit cell of the B20 lattice of MnSi. **b**, $\Delta\rho$ on a scale of $T^{3/2}$ up to about 6 K and for pressures above the critical pressure, that is, in the paramagnetic phase. From top to bottom, the pressures are 15.0, 18.1, 19.6, 25.3 and 27.5 kbar. The clear departure from a quadratic behaviour is shown in the inset for 15.0, 19.6 and 27.5 kbar.

quantum oscillatory experiments¹³ and by the low value of the residual resistivity ρ_0 , which is $0.17 \mu\Omega \text{ cm}$ for the sample used here) suggests that disorder-related effects are not expected to interfere in the experimentally relevant temperature range. The Curie temperature and the transport properties of MnSi are highly sensitive to the application of hydrostatic pressure, and may be conveniently tuned using a conventional piston and cylinder cell. In particular, T_C falls monotonically with increasing pressure and vanishes at $p_c = 14.6 \text{ kbar}$, the critical pressure of a zero-temperature phase transition. The continuous magnetic transition becomes first order when T_C is

suppressed below about 12 K, that is, for pressures above 12 kbar (refs 14–17). Crucially, there is no singularity in the magnetic susceptibility at T_C as it tends to absolute zero, and therefore p_c cannot be classified as a strict quantum critical point^{15–17}.

Here we report a high-precision study of the electrical resistivity of MnSi at pressures up to 27.5 kbar, nearly twice the critical pressure, and from room temperature down to typically 50 mK. Our main finding, shown in Figs 1 to 3, is the observation of a low-temperature electrical resistivity of the form $\rho = \rho_0 + AT^{3/2}$ in the paramagnetic phase above $p_c = 14.6 \text{ kbar}$ and up to the highest measured pressure of 27.5 kbar. Whereas the coefficient A is nearly halved over that pressure range (Fig. 1b), the exponent above p_c is weakly dependent on pressure and temperature and remains fixed at a value close to $3/2$ (Fig. 3a). The quadratic exponent normally associated with a Fermi liquid is observed only when MnSi is in a spin-polarized state, either in the zero-field ferromagnetic phase at low pressures below p_c (Figs 1a and 3a), or in the field-induced ferromagnetic state above p_c (Figs 2c and 3b). We note that the range in resistivity $\Delta\rho$ over which an exponent close to $3/2$ is observed is large compared to ρ_0 , with the ratio $\Delta\rho/\rho_0$ in excess of 17 and 14 at 15.0 and 27.5 kbar, respectively. This suggests that the $T^{3/2}$ resistivity is an intrinsic property of the pure electronic system rather than the manifestation of the effect of impurities. In addition, this anomalous resistivity is found to be robust against at least an order of magnitude increase in ρ_0 near p_c (see Methods).

The observed resistivity exponent is in conflict with the quadratic behaviour expected in Fermi-liquid theory and thought to be applicable even near p_c at low temperatures in cases where the quantum phase transition is first order. The remarkable stability of this anomalous exponent from p_c up to nearly $2p_c$, and below a critical field, is unexpected: crucially, it suggests that at low temperatures MnSi may be described in terms of a conventional spin-polarized phase in which MnSi exhibits Fermi-liquid properties and a new paramagnetic phase characterized by a non-Fermi-liquid resistivity (Fig. 3d). This is expected to be followed at sufficiently high pressures by a conventional metallic phase (not yet seen). We interpret our observation as an unusual breakdown of the scheme normally associated with ferromagnetic quantum criticality in which only a narrow crossover regime, and not a phase, is predicted. It is in this crucial sense that the present work goes beyond—but is consistent with—the previous low-temperature studies of MnSi, which were restricted to pressures only in the immediate vicinity of p_c and below¹⁰. The extensive experimental and theoretical background of this system provides us with the means to quantify the degree to which this peculiar finding is in conflict with our current understanding.

Over the past several decades, the low-temperature properties of magnetic metals such as MnSi have been described with considerable success in terms of an extension of the Fermi-liquid model in which the magnetic interactions between quasiparticles are dominant and become long range in space and time as a quantum critical point is approached. This weakly or nearly ferromagnetic Fermi-liquid (NFFL) model leads to properties that may differ strikingly from that of a normal Fermi liquid, but strictly above a characteristic temperature T^* which vanishes only at the quantum critical point itself^{18–22}. This crossover temperature T^* may be estimated in a simple way by examining the form of the relaxation frequency spectrum associated with spin fluctuations^{18,23}. For an isotropic NFFL model, the natural rate of relaxation (Γ_q) for a spontaneous spin fluctuation of wavevector q in the paramagnetic state takes the form $\Gamma_q = \theta q(\xi^{-2} + q^2)$, where $\xi = (c\chi)^{1/2}$ is the magnetic correlation length, χ is the magnetic susceptibility, and θ and c are material-specific constants²³. For the particular case of MnSi, this form for Γ_q is extensively supported by inelastic neutron scattering experiments²⁴.

The relaxation rate Γ_q is characterized by a linear term, $\theta q/\xi^2$, that represents Landau damping in a Fermi liquid, and a cubic term, θq^3 ,

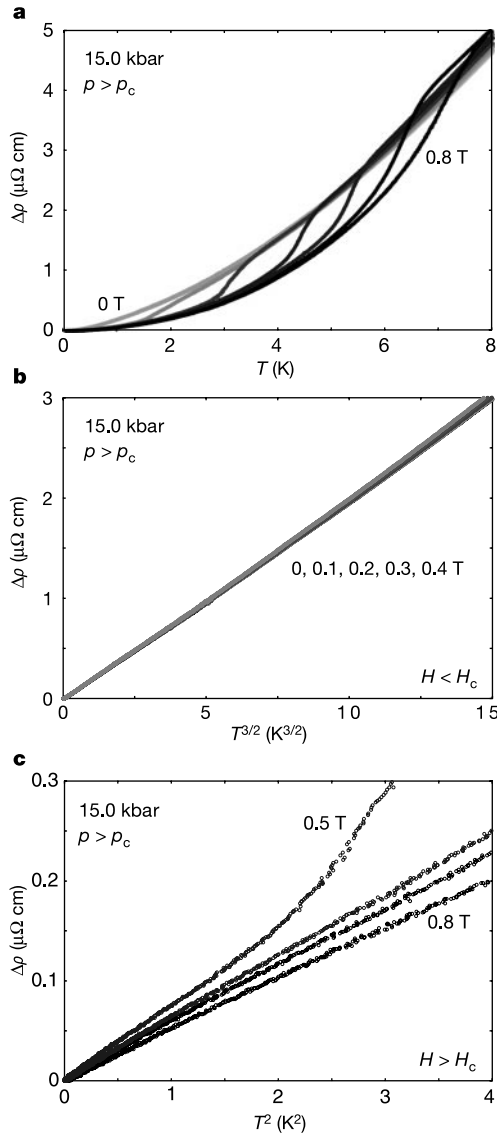


Figure 2 Dependence on temperature of the low-temperature resistivity of MnSi at 15.0 kbar and for different conditions of applied magnetic field. **a**, $\Delta\rho$ on a linear temperature scale up to 8 K. The curves shown are for 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7 and 0.8 T. For 0.4 T and below, the curves fall almost exactly on top of each other. The shoulder in the resistivity for 0.5 T and above signals a transition to a spin-polarized state at low temperatures, as known from magnetic susceptibility measurements¹⁶. The temperature associated with this transition increases with field. The zero-temperature critical field H_c associated with this field-induced ferromagnetic phase is between 0.4 and 0.5 T. **b**, $\Delta\rho$ on a scale of $T^{3/2}$ up to about 6 K and for 0, 0.1, 0.2, 0.3 and 0.4 T, that is, for $H < H_c$. **c**, $\Delta\rho$ on a scale of T^2 up to 2 K for 0.5, 0.6, 0.65 and 0.8 T, that is, for $H > H_c$. Above H_c and at low temperatures, the resistivity assumes a quadratic form similar to that seen in the zero-field ferromagnetic phase below p_c .

that becomes progressively more important as the quantum critical point is approached and ξ diverges. These two components cross at a frequency θ/ξ^3 that turns out to measure the characteristic temperature T^* (if frequencies are expressed in temperature units). The breakdown of the conventional Fermi-liquid form of the resistivity may be expected to arise only for T , and thus frequencies, well above T^* where the relaxation rate is dominated by the cubic term. As the magnetic transition near p_c is first order, ξ is not expected to be very large in MnSi and therefore T^* is not expected to be very low (see below).

T^* can be estimated from (1) inelastic neutron scattering data²⁴ at ambient pressure that define the constants θ and c , and (2) bulk susceptibility measurements¹⁵ as a function of temperature, magnetic field and pressure that, together with c , define ξ . For MnSi, we find $\theta \approx 600 \text{ K \AA}^3$ and at p_c , $\xi \approx 6 \text{ \AA}$ (ref. 23). This leads to an estimate of T^* of the order of 2 K and 10 K for pressures near p_c and $2p_c$, respectively. We stress that this estimate is in agreement with the results of direct numerical calculations of the temperature dependence of the resistivity based on the Boltzmann equation formalism¹⁸. Our estimate of T^* as a function of pressure up to $2p_c$ is represented by the dashed line in Fig. 4. So in the millikelvin range explored here (see Methods and Fig. 5), the NFFL model predicts a T^2 resistivity in essentially the entire experimental pressure and field region, in stark disagreement with our observations. This discrepancy becomes particularly difficult to ignore at the highest pressure studied here, where the non-Fermi-liquid form of the resistivity is observed down to temperatures over two orders of magnitude lower than T^* .

In an attempt to understand our observations some issues may be

raised, and here we briefly discuss two of them. First, the long-wavelength helical modulation of the ferromagnetic order is expected to lead to a modified relaxation rate of the form $\Gamma_q = \theta q(\xi^{-2} + q^2 - 2qQ)$, where Q is the wavevector of the helical modulation. But for MnSi, Q is $0.033 \text{ \AA}^{-1}$ (ref. 12) and thus much smaller than $1/\xi$ in the paramagnetic state above p_c ; that is, within the correlation length ξ the electrons are unaware of the helical modulation. This suggests that the latter leads in this case only to small corrections to the resistivity in our experimental range. In cases where Q and $1/\xi$ are comparable, however, the helical modulation may give rise to interesting and exotic behaviour (ref. 25 and M. Turlakov, personal communication).

Second, as for the effect of frozen-in disorder, the relevant length scale is the quasiparticle mean free path l . This is expected to lead to a modification of Γ_q for $q < 1/l$ and hence of the thermal properties for $T \leq \theta/(l\xi^2)$ (ref. 26). For the sample used here, this temperature scale is over three orders of magnitude smaller than T^* and therefore below the experimentally accessible range¹³. The conclusion on the ineffectiveness of disorder scattering for our pure sample is not greatly altered if we consider the relaxation rate associated with charge fluctuations rather than spin fluctuations²⁶. We also note that the low values of ρ_0 both in the ferromagnetic and paramagnetic state and the weak dependence of $\Delta\rho$ on ρ_0 suggest that the $T^{3/2}$ resistivity cannot be explained in terms of frozen-in spin disorder due to the effect of impurities as in a conventional spin glass phase.

The theory of metals has thus far been based on the concept of elementary excitations, or in the context of the present problem, on magnetic fluctuations that are well behaved and in some sense

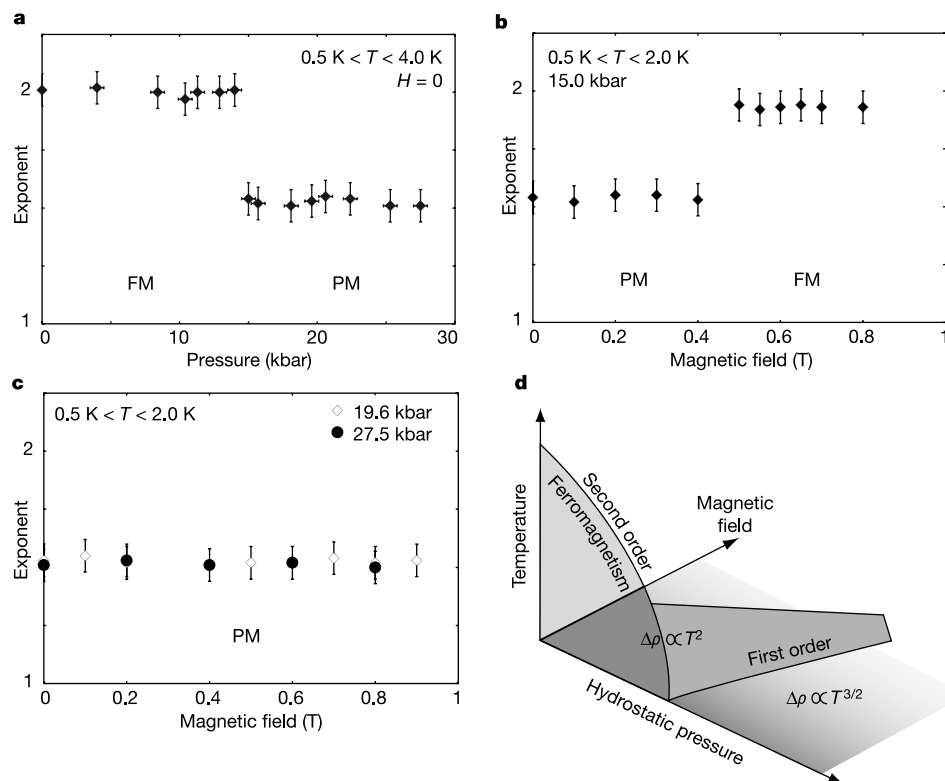


Figure 3 Low-temperature resistivity exponent as a function of pressure and magnetic field. The values are extracted from a least-squares fit to the resistivity data as a function of temperature. FM and PM denote the ferromagnetic and paramagnetic phases, respectively. **a**, Exponent as a function of pressure as observed between 0.5 and 4.0 K and in zero applied field. **b**, Exponent as a function of field as observed at 15.0 kbar between 0.5 and 2.0 K. For the data point at 0.5 T the range is 0.5 to 1.0 K because of the low transition temperature to the spin-polarized state (see Fig. 2c). **c**, Same as **b** but for

19.6 and 27.5 kbar where no spin polarization is observed in our experimental field range up to 0.9 T. Taken together, **a–c** may be seen as defining an anomalous phase in the temperature–pressure–magnetic field phase diagram with p_c and H_c for boundaries in the pressure–field plane. **d**, Illustration of the temperature–pressure–magnetic field phase diagram of MnSi supported by earlier studies^{10,14–17,30} and by our findings. The details that arise from the helical structure have been omitted.

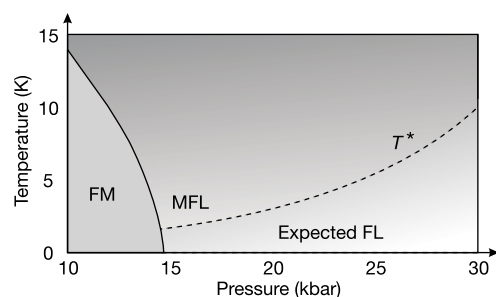


Figure 4 Illustration of the temperature-pressure phase diagram of MnSi predicted by the weakly or nearly ferromagnetic Fermi-liquid model (NFFL). In the ferromagnetic (FM) state at low temperatures and in the paramagnetic state below the crossover temperature T^* , the NFFL model reduces to the Fermi-liquid (FL) model characterized by a T^2 resistivity. Above T^* near the ferromagnetic boundary, the NFFL model reduces to the marginal Fermi-liquid (MFL) model described in our earlier work^{10,15}, which, in the ideal case of a continuous ferromagnetic transition where T^* vanishes, predicts a $T^{5/3}$ resistivity in the low-temperature limit¹⁸. In MnSi, however, the ferromagnetic transition is first order, and hence T^* remains finite at all pressures. In the ferromagnetic state and in the paramagnetic state above T^* or in high magnetic fields, the NFFL model provides a consistent quantitative description of the known properties of MnSi. Below T^* in the paramagnetic phase, however, our measurements conflict in a profound and crucial way with this phase diagram, in that no resistivity consistent with a Fermi-liquid description is observed anywhere below T^* , from p_c up to 27.5 kbar and down to 50 mK, that is, up to two orders of magnitude below T^* . We note that in calculating T^* we have assumed the parameters θ and c to be constants independent of temperature, magnetic field and pressure. We find, however, that within the NFFL model, no simple or plausible variation of these parameters can consistently account for the overall behaviour of the resistivity and of the magnetic susceptibility over the entire temperature, field and pressure ranges investigated.

smooth¹. However, on the border of a first-order phase transition one might expect, in addition to these smooth fluctuations, the emergence of sudden large-amplitude fluctuations representing the spontaneous formation of droplets of local order. These highly anharmonic quantum and thermal fluctuations may not be described simply in terms of elementary excitations. Normally, such anharmonic fluctuations are frozen out by a high-energy barrier between the paramagnetic and locally ferromagnetic state, and lead to the well-known supercooling phenomenon and hysteresis associated with conventional first-order transitions²⁷. In materials such as MnSi, however, the first-order transition may be weak enough to permit anharmonic fluctuations to play a new role in the quantum statistical mechanics of a metal at low temperature (T. V. Ramakrishnan and C. M. Varma, personal communication)²².

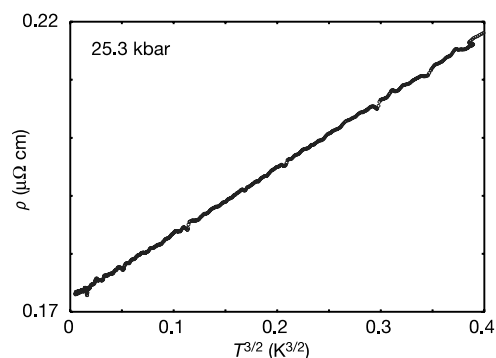


Figure 5 Dependence on temperature of the electrical resistivity of MnSi at subkelvin temperatures. In this range and for high-purity specimens, we have observed an exponent between 1.5 and 1.7. For this particular trace at 25.3 kbar and measured with a dilution refrigerator, ρ varies as $T^{3/2}$.

Evidence of anomalous behaviour on the border of first-order phase transitions that may support this view has recently been reported in other contexts²⁸. The relevance of our findings in MnSi to other d -metal ferromagnets is the subject of several independent investigations (C. Pfeleiderer, J. Flouquet, P. Niklowitz and D. Jaccard, personal communications). □

Methods

Samples

The single-crystal samples were grown by radio-frequency induction melting of zone-refined high-purity starting materials on a water-cooled copper crucible in an ultrahigh vacuum. The resulting ingots were annealed for two weeks near the melting point. Single-crystal samples were oriented by Laue X-ray diffraction, and cut to typically $2.0 \times 0.5 \times 0.5 \text{ mm}^3$ using low-power spark erosion. Four electrical contacts were made immediately after etching the samples for 20 min in a solution of glycerol:HF:HCl:HNO₃ 40:25:25:10. At ambient pressure, the elastic mean free path found in quantum oscillations measurements is in excess of 5,000 Å, in agreement with $\rho_0 = 0.17 \mu\Omega \text{ cm}$.

Experimental procedure

The high-pressure experiment was performed using a 4-mm piston-and-cylinder clamp cell made exclusively of non-magnetic materials²⁹. To achieve a high degree of hydrostaticity we employ as pressure medium an equal mix of *iso*-pentane and *n*-pentane known to remain liquid at room temperature in the experimentally relevant pressure range. The pressure cell is gold-plated for optimum thermal contact and to reduce the absorption of thermal radiation.

The use of ultra-low-noise a.c. detection based on cryogenically cooled electronics eliminates common problems related to, for example, thermal noise or drift effects, and allows measurements at very low excitation levels, even in the millikelvin range.

Measurements were performed and repeated on three different experimental systems, one conventional dilution refrigerator with a base temperature of 15 mK, and two demagnetization refrigerators reaching typically 50 mK. The slow sweep rate, of the order of 1 mK min^{-1} at low temperatures, ensures thermal equilibrium between the samples and the thermometry.

We stress that in the present work measurements were systematically performed on a number of samples and on all three experimental systems. Despite focusing here on the findings related to our purest sample, the observations reported here were also seen near p_c in samples of varying quality with residual resistivities as high as $2.4 \mu\Omega \text{ cm}$.

Received 30 May; accepted 30 July 2003; doi:10.1038/nature01968.

- Landau, L. D. *Collected Papers* (ed. Ter Haas, D.) Ch. 90, 91 (Pergamon, Oxford, 1965).
- Murata, K. K. & Doniach, S. Theory of magnetic fluctuations in itinerant ferromagnets. *Phys. Rev. Lett.* **29**, 285–288 (1972).
- Moriya, T. & Kawabata, A. Effect of spin fluctuations in itinerant electron ferromagnetism. *J. Phys. Soc. Jpn* **34**, 639–651 (1973).
- Ramakrishnan, T. V. Microscopic theory of spin fluctuations in itinerant electron ferromagnets. *Phys. Rev. B* **10**, 4014–4024 (1974).
- Hertz, J. A. Quantum critical phenomena. *Phys. Rev. B* **14**, 1165–1184 (1976).
- Laughlin, R. B. *et al.* The quantum criticality conundrum. *Adv. Phys.* **50**, 361–365 (2001).
- Anderson, P. W. In praise of unstable fixed points: The way things actually work. *Physica B* **318**, 28–32 (2002).
- Schröder, A. *et al.*, Onset of antiferromagnetism in heavy-fermion metals. *Nature* **407**, 351–355 (2000).
- Millis, A. J. Whither correlated electron theory? *Physica B* **312–313**, 1–6 (2002).
- Pfeleiderer, C., Julian, S. R. & Lonzarich, G. G. Non-Fermi-liquid nature of the normal state of itinerant-electron ferromagnets. *Nature* **414**, 427–430 (2001).
- Bloch, D., Voiron, J., Jaccarino, V. & Wernick, J. H. The high field-high pressure magnetic properties of MnSi. *Phys. Lett.* **51**, 259–291 (1975).
- Lebech, B. in *Recent Advances in Magnetism of Transition Metal Compounds* eds Kotani, A. & Suzuki, N.) 167–178 (World Scientific, Singapore, 1993).
- Taillefer, L., Lonzarich, G. G. & Strange, P. The band magnetism of MnSi. *J. Magn. Magn. Mater.* **54–57**, 957–958 (1986).
- Thompson, J. D., Fisk, Z. & Lonzarich, G. G. Perspective on heavy-electron and Kondo-lattice systems from high pressure studies. *Physica B* **161**, 317–323 (1989).
- Pfeleiderer, C., McMullan, G. J., Julian, S. R. & Lonzarich, G. G. Magnetic quantum phase transition in MnSi under hydrostatic pressure. *Phys. Rev. B* **55**, 8330–8338 (1997).
- Thessieu, C. *et al.* Field dependence of the magnetic quantum phase transition in MnSi. *J. Phys. Condens. Matter* **9**, 6677–6687 (1997).
- Thessieu, C., Ishida, K., Kitaoka, Y., Asayama, K. & Lapertot, G. Pressure effect on MnSi: An NMR study. *J. Magn. Magn. Mater.* **177–181**, 609–610 (1998).
- Moriya, T. *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin, 1985).
- Millis, A. J. Effect of a non-zero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48**, 7183–7196 (1993).
- Pfeleiderer, C. Non-Fermi liquid puzzle of MnSi at high pressure *Physica B* **328** (1–2), 100–104 (2003).
- Belitz, D., Kirkpatrick, T. R., Narayanan, R. & Vojta, T. Transport anomalies and marginal-Fermi-liquid effects at a quantum critical point. *Phys. Rev. Lett.* **85**, 4602–4605 (2000).
- Varma, C. M., Nussinov, Z. & van Saarloos, W. Singular or non-Fermi liquids. *Phys. Lett.* **361**, 267–417 (2002).
- Lonzarich, G. G. & Taillefer, L. Effect of spin fluctuations on the magnetic equation of state of ferromagnetic or nearly ferromagnetic metals. *J. Phys. C* **18**, 4339–4371 (1985).
- Ishikawa, Y., Noda, Y., Uemura, Y. J., Majkrzak, C. F. & Shirane, G. Paramagnetic spin fluctuations in the weak itinerant-electron ferromagnet MnSi. *Phys. Rev. B* **31**, 5884–5893 (1985).

25. Vojta, T. & Sknepnek, R. Quantum phase transition of itinerant helimagnets. *Phys. Rev. B* **64**, 052404 (2001).
26. Altshuler, B. L. & Aronov, A. G. in *Electron-Electron Interactions in Disordered Systems* (eds Pollak, M. & Efros, A. L.) 1–153 (Modern Problems in Condensed Matter Sciences, North Holland, Amsterdam, 1985).
27. Langer, J. S. Statistical theory of the decay of metastable states. *Ann. Phys.* **54**, 258–275 (1969).
28. Rivadulla, E., Zhou, J.-S. & Goodenough, J. B. Electron scattering near an itinerant to localized electronic transition. *Phys. Rev. B* **67**, 165110 (2003).
29. Walker, I. R. Nonmagnetic piston-cylinder pressure cell for use at 35 kbar and above. *Rev. Sci. Instrum.* **70**, 3402–3412 (1999).
30. Yamada, H. & Terao, K. Itinerant-electron metamagnetism of MnSi at high pressure. *Phys. Rev. B* **59**, 9342–9347 (1999).
31. Custers, J. *et al.*, The break-up of heavy electrons at a quantum critical point. *Nature* **424**, 524–527 (2003).

Acknowledgements We thank the following for discussions; B. Altshuler, S. Barakat, S. Brown, P. Coleman, J. Flouquet, R. K. W. Haselwimmer, D. Khmelnitskii, A. J. Millis, P. Monthoux, P. Niklowitz, C. Pfleiderer, T. V. Ramakrishnan, S. S. Saxena, B. Simon, M. Turlakov and C. M. Varma. The work was supported by the UK EPSRC and the EU FERLIN programme. N.D.-L. acknowledges support from FCAR of Quebec, NSERC of Canada, and Trinity College and Peterhouse, Cambridge University.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.D.-L. (nd223@cam.ac.uk).

Temperature-induced valence transition and associated lattice collapse in samarium fulleride

J. Arvanitidis¹, Konstantinos Papagelis¹, Serena Margadonna², Kosmas Prassides¹ & Andrew N. Fitch³

¹Department of Chemistry, University of Sussex, Brighton BN1 9QJ, UK

²Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK

³European Synchrotron Radiation Facility, 38042 Grenoble, France

The different degrees of freedom of a given system are usually independent of each other but can in some materials be strongly coupled, giving rise to phase equilibria sensitively susceptible to external perturbations. Such systems often exhibit unusual physical properties that are difficult to treat theoretically, as exemplified by strongly correlated electron systems such as intermediate-valence rare-earth heavy fermions and Kondo insulators, colossal magnetoresistive manganites and high-transition temperature (high- T_c) copper oxide superconductors. Metal fulleride salts¹—metal intercalation compounds of C_{60} —and materials based on rare-earth metals also exhibit strong electronic correlations. Rare-earth fullerides thus constitute a particularly intriguing system—they contain highly correlated cation (rare-earth) and anion (C_{60}) sublattices. Here we show, using high-resolution synchrotron X-ray diffraction and magnetic susceptibility measurements, that cooling the rare-earth fulleride $Sm_{2.75}C_{60}$ induces an isosymmetric phase transition near 32 K, accompanied by a dramatic isotropic volume increase and a samarium valence transition from $(2 + \epsilon)^+$ to nearly 2^+ . The negative thermal expansion—heating from 4.2 to 32 K leads to contraction rather than expansion—occurs at a rate about 40 times larger than in ternary metal oxides typically exhibiting such behaviour². We attribute the large negative thermal expansion, unprecedented in fullerene or other molecular systems, to a quasi-continuous valence transition from $Sm^{2+\epsilon}$ towards the smaller $Sm^{(2+\epsilon)+}$, analogous to the valence or configuration transitions encountered in intermediate-valence Kondo insulators like SmS (ref. 3).

The transition between superconducting and antiferromagnetic insulating behaviour of C_{60} -based fullerides is sensitively controlled by the ratio (U/W) between the on-site Coulomb repulsion energy, U , and the conduction bandwidth, W (ref. 1). Whereas intense efforts in the last decade have exhaustively mapped the electronic and superconducting properties of intercalated alkali and alkaline-earth fullerides^{4–6}, little is known about the electronic properties of other metal fullerides, especially those of rare-earth elements, Ln, which present the possibility of strong coupling between two electronically active sublattices, $\pi(C_{60})$ and $4f,5d(Ln)$. The most prominent examples of rare-earth fullerides are the $Yb_{2.75}C_{60}$ ($T_c = 6$ K)⁷ and $Sm_{2.75}C_{60}$ ($T_c = 8$ K)⁸ superconductors, whose complex orthorhombic structures are related to those of primitive cubic sodium fullerides⁹ and arise from the long-range ordered arrangement of tetrahedral rare-earth metal vacancies, accompanied by orientational ordering of C_{60} units about local three-fold symmetry rotational axes⁷.

First, we found no evidence in $Sm_{2.75}C_{60}$ for a transition to a superconducting state down to 2 K in contrast to ref. 8 and in agreement with other explorations of the $Sm-C_{60}$ phase field¹⁰. We then probed the temperature evolution of the structural properties of the $Sm_{2.75}C_{60}$ fulleride by high-resolution synchrotron X-ray diffraction, and examined the diffraction profiles collected over a restricted angular range of 3.3° to 25° (d -spacing, 13.9 to 1.85 Å) at various temperatures between 4.2 and 295 K. Inspection of the diffraction profiles readily reveals the appearance of superlattice peaks at low angles that index to the enlarged orthorhombic unit cell (space group $Pcab$) proposed before^{7,11}. However, a striking feature of the data is that the diffraction peaks at low temperatures continuously shift to higher angles on heating, implying an anomalous structural behaviour whereby the material rapidly contracts as the temperature increases above 4.2 K (Fig. 1). The trend is reversed above 32 K and normal behaviour is restored with the lattice smoothly expanding on heating to 295 K. Besides the shifts in peak positions, no changes in relative peak intensities are apparent in the diffraction profiles throughout the whole temperature range, implying the absence of a phase transition to a structure with

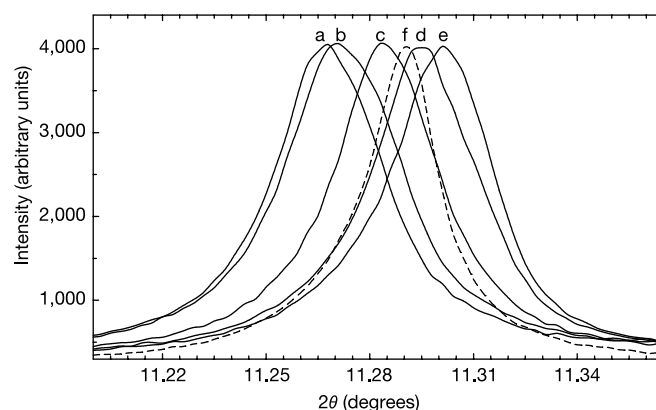


Figure 1 Selected region of the diffraction profile of $Sm_{2.75}C_{60}$ showing the temperature evolution of the (444) Bragg reflection ($\lambda = 0.79980$ Å). Trace a, 4.2 K; b, 10 K; c, 16 K; d, 21 K; e, 32 K; and f, 200 K. The peak shifts monotonically to higher angles (lattice contraction) on heating from 4.2 to 32 K and then to lower angles (lattice expansion) on further heating to room temperature. Its width remains essentially constant in the temperature range in which the negative thermal expansion is observed. Then it sharpens continuously on heating to 150 K (by $\sim 45\%$) with little further change to room temperature. Such behaviour could reflect the presence of a structural phase transition undetectable even with the ultrahigh resolution of the present measurements. More likely it could be the signature of precursor effects to the valence transition and the existence of local structural inhomogeneities accompanying the transformation.

different crystal symmetry detectable with the ultrahigh resolution of the present data. Extraction of reliable lattice constants was performed with the LeBail pattern decomposition technique using the same space group $Pcab$ at all temperatures (at 4.2 K: $a = 28.21778(7)$ Å, $b = 28.25426(7)$ Å, $c = 28.18970(7)$ Å; agreement factors, $R_{wp} = 2.11\%$, $R_{exp} = 1.30\%$). Figure 2 shows the temperature evolution of the lattice constants and of the corresponding unit cell volume. Initially the lattice contracts slowly on heating above 4.2 K. Then the rate of contraction increases to $d\ln V/dT \approx -400$ p.p.m. K^{-1} , resulting in an overall volume decrease of 0.84% between 4.2 and 32 K (Fig. 2 inset). The sign of the thermal expansivity then changes and the lattice expands on heating to 295 K at a rate of approximately $+20$ p.p.m. K^{-1} , comparable to that typically encountered in other metal fulleride salts (for example, for $K_3Ba_3C_{60}$, $d\ln V/dT \approx +30$ p.p.m. K^{-1})¹². Despite this, the lattice dimensions of $Sm_{2.75}C_{60}$ at room temperature remain smaller than those at 4.2 K.

Refinement of the structure of $Sm_{2.75}C_{60}$ was then attempted by Rietveld analysis of the diffraction data (Supplementary Fig. 1S) collected over an extended angular range of 3.4° to 50° (d -spacing, 13.5 to 0.95 Å) at 5 and 34 K. A perspective view of the refined structure of $Sm_{2.75}C_{60}$ is shown in Fig. 3. The final average Sm–C

distances are 2.91 and 2.94 Å for the tetrahedral and octahedral sites, respectively. The results of the final refinement at 5 K are shown in Supplementary Fig. 1S ($R_{wp} = 3.67\%$, $R_{exp} = 1.15\%$; $a = 28.2025(2)$ Å, $b = 28.2397(2)$ Å, $c = 28.1632(2)$ Å) with the fitted parameters summarized in Supplementary Tables 1S and 2S. The lattice dimensions derived from the analysis of these diffraction data at 5 K are somewhat smaller than those extracted at the same temperature during the stepwise cooling protocol, implying the existence of hysteretic effects associated with the transition. No discernible differences are evident between the refinements at 5 and 34 K, except for the lattice contraction effect, thereby discarding the possibility of the existence of a structural phase transition accompanied by a crystal symmetry change.

The most prominent point arising from the results of the present structural refinements is the remarkable lattice response at low temperatures. The negative thermal expansion and quasi-continuous collapse of the structure of $Sm_{2.75}C_{60}$ on heating are not accompanied by a change in crystal symmetry and the full space group is preserved. The origin of negative thermal expansion behaviour over broad temperature ranges in oxides like ZrW_2O_8 (ref. 2) has been associated with the presence of low-energy highly anharmonic transverse thermal vibrations, which tend to contract the lattice. Although anomalies in the low-energy phonon spectrum may well occur in $Sm_{2.75}C_{60}$ through the transition, possibly associated with strong coupling between Sm and C_{60} motion, we favour an explanation of the effect based on the fragility of the valence states of Sm and its tendency to exhibit intermediate valence or configuration characteristics. We also recall that the unit cell size of metal fullerides is sensitively controlled by the size of the ions residing in the small tetrahedral interstices and ions with a size, r , larger than that of the tetrahedral hole, r_{tetr} are accommodated with substantial expansion of the lattice. In $Sm_{2.75}C_{60}$, the tetrahedral hole has a radius of 1.12 Å, straddling the values of the ionic radii of Sm^{2+} (1.14 Å) and Sm^{3+} (0.96 Å). A valence transition of Sm from nearly $2+$ ($r > r_{tetr}$) towards an intermediate value of $(2 + \epsilon)+$ with smaller ionic size ($r < r_{tetr}$), induced on heating, will thus have

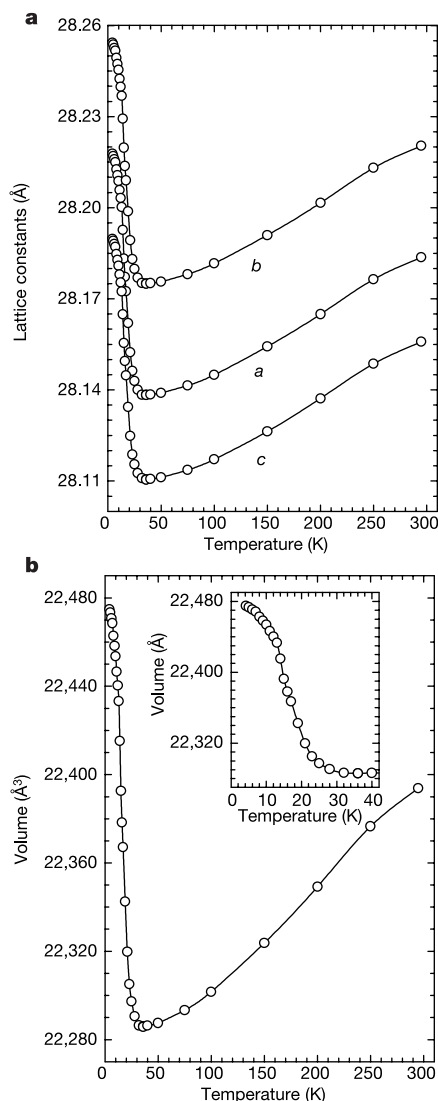


Figure 2 Temperature evolution of structural parameters of $Sm_{2.75}C_{60}$. **a**, The orthorhombic lattice constants; **b**, the unit cell volume. The inset in **b** shows an expanded view of the volume dependence on temperature in the range between 4.2 and 40 K.

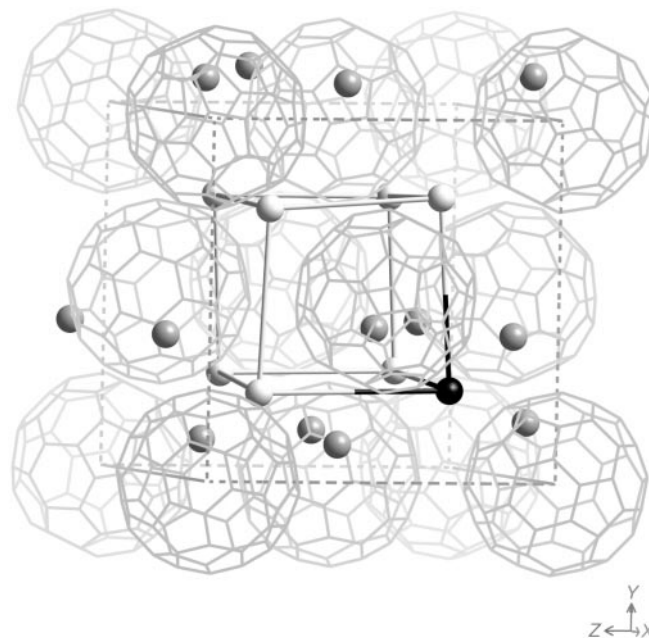


Figure 3 Building block of the orthorhombic superstructure of $Sm_{2.75}C_{60}$, which can be obtained by doubling along all three lattice directions. Distorted octahedral and tetrahedral samarium cations are depicted as dark- and light-grey spheres, respectively. The tetrahedral samarium defect is shown as a black sphere. The C_{60} units are optimally rotated about local three-fold symmetry axes, as described in the Methods section.

a profound effect on the lattice size and is consistent with the observed dramatic volume decrease.

Such mixed valence behaviour has been well documented in rare-earth compounds like the samarium monochalcogenides. For instance, SmS (NaCl-type structure) undergoes an abrupt (first-order) catastrophic transition from a semiconducting black phase to a metallic gold one at 0.65 GPa at ambient temperature³. In contrast to SmS, the transformation is continuous in SmSe and occurs over a broad pressure range (0–5 GPa)³. In both cases, the transitions are isosymmetric and are characterized by a large volume collapse consistent with the conversion of Sm^{2+} to Sm^{3+} . A charge (valence) fluctuation model, $2 + (4f^6 5d^0) \leftrightarrow 3 + (4f^5 5d^1)$, has been employed to describe the physics of these systems with the valence transition controlled by the Kondo temperature, which reflects the strength of the hybridization between the $4f$ and the $5d$ conduction electrons. The transition pressure can be successfully lowered by alloying SmS with trivalent rare-earth metals, Ln, to give $\text{Sm}_{1-x}\text{Ln}_x\text{S}$ compositions. In such cases and for selected values of the dopant level x , the collapsed intermediate valence ($\text{Sm}^{(2+\epsilon)+}$) phases can be stabilized at ambient pressure, whereby upon cooling they exhibit either continuous or discontinuous valence transitions accompanied by expansion of the lattice dimensions¹³, much like what is observed here for $\text{Sm}_{2.75}\text{C}_{60}$.

Figure 4 shows the results of temperature-dependent magnetic measurements at 1 T for $\text{Sm}_{2.75}\text{C}_{60}$. The diagram also includes the calculated magnetic susceptibilities of the free Sm^{2+} and Sm^{3+} ions. For Sm^{2+} ($4f^6$, 7F_0), the Curie contribution of the ground state to the magnetic susceptibility is explicitly zero and there is only a van Vleck term arising through the mixing of the $J = 0$ ground and the $J = 1$ excited state (energy separation, 420 K). The susceptibility of Sm^{3+} ($4f^5$, $^6H_{5/2}$) is obtained by summing the corresponding Curie and van Vleck (energy separation between $J = 5/2$ ground and $J = 7/2$ excited state is 348 K) contributions. The samarium valence in intermediate valence compounds may be typically derived from magnetic susceptibility data at sufficiently high temperatures; at such temperatures the measuring scale (kT) is faster than the fluctuation frequency (ω_{sf}) of the magnetic moment between the values of the two valence configurations¹⁴, and the magnetic susceptibility can then be expressed as a linear combination of the Sm^{2+} and Sm^{3+} contributions, $\chi = (1 - \epsilon)\chi(\text{Sm}^{2+}) + \epsilon\chi(\text{Sm}^{3+})$, where $(2 + \epsilon)$ is approximately the average Sm valence. This

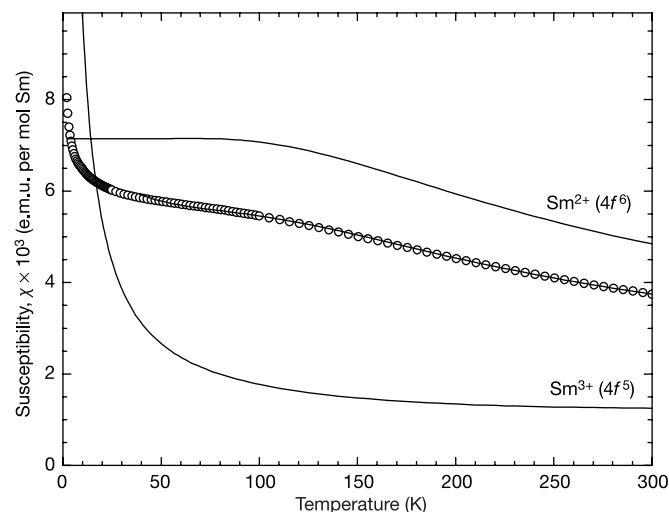


Figure 4 Temperature dependence of magnetic susceptibilities. Open symbols show the measured magnetic susceptibility of $\text{Sm}_{2.75}\text{C}_{60}$ in a field $H = 1$ T. The calculated magnetic susceptibilities (solid lines) of the free Sm^{2+} and Sm^{3+} ions are also included. The line through the experimental points is the weighted average of the susceptibility calculated assuming 70% and 30% contributions from Sm^{2+} and Sm^{3+} , respectively.

relationship does not hold below some characteristic temperature, T_{sf} . Along the same lines, we find that the room-temperature value of the magnetic susceptibility of $\text{Sm}_{2.75}\text{C}_{60}$ ($\chi_{\text{RT}} = 3.75 \times 10^{-3}$ e.m.u. per mol Sm) straddles those calculated for the free Sm^{2+} and Sm^{3+} ions, thereby providing an estimate of the average Sm valence as approximately equal to 2.3+. We also include in Fig. 4 the temperature dependence of the calculated average susceptibility assuming 70% and 30% contributions from Sm^{2+} and Sm^{3+} , respectively. This follows the experimental data very well down to about 40 K, just above the onset of the valence transition. It then diverges significantly, as the suppressed values of the measured magnetic susceptibility now reflect the transition to a Sm valence closer to 2+.

The data presented here indicate that the behaviour of the $\text{Sm}_{2.75}\text{C}_{60}$ fulleride—which shows a quasi-continuous isosymmetric phase transition below 32 K to a significantly expanded lattice—resembles the behaviour of intermediate valence compounds of the rare-earths, and points towards an electronic origin of the transition, with strong coupling between the charge (valence), lattice and spin degrees of freedom. They also open the way for the isolation and study of new families of molecular-based strongly correlated Kondo and heavy-fermion systems in which both cation and anion sublattices are electronically active. It would be interesting to explore in detail the consequences of the valence (configuration) transition for the electronic and conducting properties (especially as the C_{60} sublattice can also support superconductivity), to search for pressure-induced effects that are highly anticipated, and to characterize the dynamics of the interconfigurational fluctuations. □

Methods

Sample preparation

$\text{Sm}_{2.75}\text{C}_{60}$ samples were prepared by reaction of stoichiometric quantities of C_{60} and Sm, pressed into pellets and contained in a sealed tantalum cell inside an evacuated quartz tube at 575 °C for 3 days with one intermediate grinding.

Synchrotron X-ray diffraction and magnetization

For the synchrotron X-ray diffraction measurements, a sample was sealed in a thin-walled glass capillary 0.5 mm in diameter. With the sample inside a continuous-flow cryostat, synchrotron X-ray powder diffraction data ($\lambda = 0.79980$ Å, $2\theta = 3.4^\circ$ to 50°) were collected at 5 and 34 K in continuous scanning mode with the high-resolution multianalyser powder diffractometer on beamline ID31 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Diffraction profiles were also recorded on heating at numerous temperatures between 4.2 and 295 K over a shorter angular range ($2\theta = 3.3^\circ$ to 25°). Data analysis was performed with the GSAS suite of Rietveld analysis programs. Magnetization measurements with a Quantum Design MPMS5 SQUID (superconducting quantum interference device) susceptometer on numerous samples (mass, ~50 mg) sealed in quartz tubes revealed no evidence of superconductivity. Total susceptibilities were obtained for the $\text{Sm}_{2.75}\text{C}_{60}$ samples in the temperature range 2 to 300 K in a field of 1 T after correcting for the diamagnetic core contributions.

Rietveld analysis of the synchrotron X-ray powder diffraction data

As there is a remarkable correspondence of the reflection intensities in the diffraction profile of $\text{Sm}_{2.75}\text{C}_{60}$ with those observed in $\text{Yb}_{2.75}\text{C}_{60}$, the Rietveld refinement was initiated in the orthorhombic space group $Pcab$ with the samarium cations occupying off-centred tetrahedral and octahedral interstitial sites⁷. One out of every eight tetrahedral sites was initially empty, with long-range ordering of these tetrahedral vacancies resulting in a unit cell with dimensions twice as large as those of the commonly encountered face-centred cubic fulleride structures. The five symmetry-inequivalent C_{60} molecules present in the unit cell ($\text{C}_{60}(1)$ at (000), $\text{C}_{60}(21)$ at (0,1/4,1/4), $\text{C}_{60}(22)$ at (1/4,0,1/4), $\text{C}_{60}(23)$ at (1/4,1/4,0) and $\text{C}_{60}(3)$ at (1/2,1/2,1/2)) were rotated anticlockwise by 37.5° about local symmetry axes ($[111]$, $[1\bar{1}\bar{1}]$, $[\bar{1}\bar{1}1]$ and $[\bar{1}1\bar{1}]$, respectively) following ref. 7. Stable refinements were quickly achieved, but even though agreement between observed and calculated profiles became very good (agreement factors: $R_{\text{wp}} = 4.19\%$, $R_{\text{exp}} = 1.15\%$), there were signs of systematic differences between model and experimental intensities. We thus decided to allow partial occupation of the tetrahedral vacancy by Sm. The Rietveld refinement immediately improved significantly with the fraction of the Sm defect converging to ~25%. The obtained structural model is characterized by large displacements of the cations from the centres of the octahedral sites (~2.3 Å) and smaller displacements from the centres of the tetrahedral sites (~0.4 Å). However, the crystallographically distinct tetrahedral Sm ions in the unit cell do not have identical coordination environments, and careful examination of the metal-carbon distances revealed the presence of unphysically short contacts (2.52(3) Å) between Sm(3) located off the centre of the (3/8,3/8,3/8) tetrahedral hole and $\text{C}_{60}(3)$. Noting that the $\text{C}_{60}(3)$ unit is

surrounded by the maximum possible number of 12 near-neighbour Sm cations¹⁵, we attempted to relieve the steric strain by exploring the suitability of alternative orientations. A satisfactory solution was found when C₆₀(3) was rotated anticlockwise by 37.5° about the local [111] symmetry axis. This led to both an improved agreement between observed and calculated profiles and to physically meaningful shortest Sm(3)–C contacts (2.78(3) Å). The superior quality, very high resolution and extended angular range of the present data also allowed us to discard the possibility that the observed superstructure in Sm_{2.75}C₆₀ retains primitive cubic symmetry (space group *Pa*3̄). Rietveld refinements with this space group always led to significantly worse agreement factors.

Received 14 May; accepted 12 August 2003; doi:10.1038/nature01994.

- Gunnarsson, O. Superconductivity in fullerenes. *Rev. Mod. Phys.* **69**, 575–606 (1997).
- Mary, T. A., Evans, J. S. O., Vogt, T. & Sleight, A. W. Negative thermal expansion from 0.3 to 1050 K in ZrW₂O₈. *Science* **272**, 90–92 (1996).
- Jayaraman, A., Narayana, V., Bucher, E. & Maines, R. G. Continuous and discontinuous semiconductor-metal transition in samarium monochalcogenides under pressure. *Phys. Rev. Lett.* **25**, 1430–1433 (1970).
- Rosselinsky, M. J. Recent developments in the chemistry and physics of metal fullerenes. *Chem. Mater.* **10**, 2665–2685 (1998).
- Margadonna, S. & Prassides, K. Recent advances in fullerene superconductivity. *J. Solid State Chem.* **168**, 639–652 (2002).
- Iwasa, Y. & Takenobu, T. Superconductivity, Mott–Hubbard states, and molecular orbital order in intercalated fullerenes. *J. Phys. Condens. Matter* **15**, R495–R519 (2003).
- Özdas, E. *et al.* Superconductivity and cation-vacancy ordering in the rare-earth fulleride Yb_{2.75}C₆₀. *Nature* **375**, 126–129 (1995).
- Chen, X. H. & Roth, G. Superconductivity at 8 K in samarium-doped C₆₀. *Phys. Rev. B* **52**, 15534–15536 (1995).
- Prassides, K. *et al.* Crystal structure, bonding, and phase transition of the superconducting Na₂CsC₆₀ fulleride. *Science* **263**, 950–954 (1994).
- Claves, D. & Hamwi, A. The polymorphism of samarium fullerenes. *Solid State Comm.* **113**, 357–362 (2000).
- Sun, Z., Chen, X. H., Takenobu, T. & Iwasa, Y. Structure and Raman behaviour of Sm_{2.75}C₆₀. *J. Phys. Condens. Matter* **12**, 8919–8926 (2000).
- Margadonna, S. *et al.* Crystal structure of superconducting K₃Ba₃C₆₀: A combined synchrotron X-ray and neutron diffraction study. *Chem. Mater.* **12**, 2736–2740 (2000).
- Jayaraman, A., Bucher, E., Dernier, P. D. & Longinotti, L. D. Temperature-induced explosive first-order electronic phase transition in Gd-doped SmS. *Phys. Rev. Lett.* **31**, 700–703 (1973).
- Maple, M. B. & Wohlleben, D. Nonmagnetic 4f shell in the high-pressure phase of SmS. *Phys. Rev. Lett.* **27**, 511–514 (1971).
- Rabe, K. M. & Citrin, P. H. Vacancy-induced structural relaxations in Yb_{2.75}C₆₀. *Phys. Rev. B* **58**, R551–R554 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the ESRF for synchrotron X-ray beamtime, the Marie Curie Fellowship programme of the European Union “Improving the Human Research Potential” for support (J.A., K.P.) and the Royal Society for a Dorothy Hodgkin Research Fellowship (S.M.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.P. (K.Prassides@sussex.ac.uk) or S.M. (sm413@cam.ac.uk).

Understanding and tuning the epitaxy of large aromatic adsorbates by molecular design

M. Eremtchenko¹, J. A. Schaefer¹ & F. S. Tautz²

¹Institut für Physik und Zentrum für Mikro- und Nanotechnologien, Technische Universität Ilmenau, PO Box 100565, 98684 Ilmenau, Germany

²School of Engineering and Science, International University Bremen, PO Box 750561, 28725 Bremen, Germany

If the rich functionality of organic molecules is to be exploited in devices such as light-emitting diodes or field-effect transistors^{1–7}, interface properties of organic materials with various (metallic and insulating) substrates must be tailored carefully^{7–10}. In many cases, this calls for well-ordered interfaces. Organic epitaxy^{11–13}—that is, the growth of molecular films with a commensurate

structural relationship to their crystalline substrates—relies on successful recognition of preferred epitaxial sites. For some large π -conjugated molecules (‘molecular platelets’) this works surprisingly well^{14,15}, even if the substrate exhibits no template structure into which the molecules can lock^{13,15,16}. Here we present an explanation for site recognition in non-templated organic epitaxy, and thus resolve a long-standing puzzle¹¹. We propose that this form of site recognition relies on the existence of a local molecular reaction centre in the extended π -electron system of the molecule. Its activity can be controlled by appropriate side groups and—in a certain regime—may also be probed by molecularly sensitized scanning tunnelling microscopy. Our results open the possibility of engineering epitaxial interfaces, as well as other interfacial nanostructures for which specific site recognition is essential.

The case of 3,4,9,10-perylene-tetracarboxylicacid-dianhydride (PTCDA; Fig. 1c) on Ag(111)¹³ is one of the most prominent examples of organic epitaxy^{14–17}. Here we distinguish between true commensurate epitaxy and so-called quasi-epitaxy^{18,19}. The latter—with its incommensurate, coincident or so-called point-on-line lattice registries—is most commonly observed for large organic molecules (for example, polyacenes and their derivatives) on substrates as diverse as passivated semiconductor surfaces²⁰, close-packed metal surfaces²¹, graphite²² and glass²³. It has been pointed out that molecules which form layered van der Waals bonded crystals in their bulk states as a rule allow quasi-epitaxy on inorganic substrates¹¹. The reason can be found in their large interlayer (shear-) elastic compliance, resulting also in a very shallow minimum (with respect to relative translations or rotations) of the interaction potential between molecular layer and substrate. Yet there are notable but little-understood exceptions to this rule. If PTCDA is deposited onto Ag(111) substrates by physical vapour deposition under ultrahigh vacuum, it forms almost perfectly ordered commensurate overlayers¹³. Figure 1a shows a scanning tunnelling microscopy (STM) image of such an epitaxial monolayer, exhibiting the characteristic ‘herringbone’ pattern known from the bulk crystal structure of PTCDA²⁴. This arrangement results from the minimization of electrostatic interaction energy originating in the quadrupolar field associated with the molecule.

In the STM image displayed in Fig. 1a, two sets of molecules are imaged with different brightnesses. This contrast difference has been explained by a site-specific chemical bonding¹³. The absence

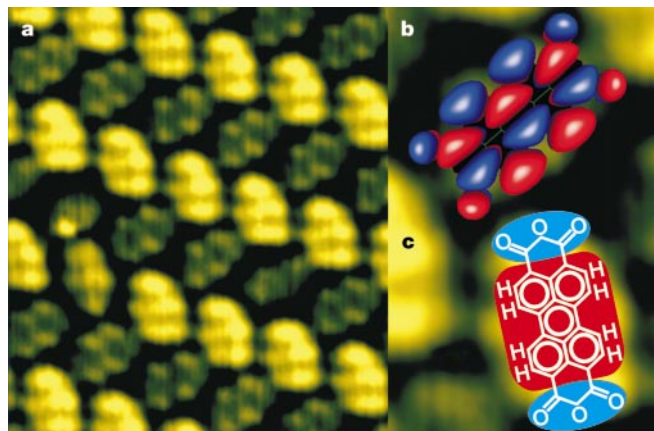


Figure 1 Organic epitaxy of PTCDA on the Ag(111) surface. **a**, STM image of PTCDA ‘herringbone’ layer on Ag(111), image area $7 \times 7 \text{ nm}^2$ (tunnelling parameters: $I_T = 0.46 \text{ nA}$, $U_S = -2.3 \text{ V}$). **b**, Calculated probability density of the LUMO²⁸ of PTCDA. **c**, Chemical structure of PTCDA, with schematic indication of intramolecular charge distribution (red shows positive, and blue shows negative, partial charges).

of any additional contrast modulation indicates an epitaxial monolayer: all molecules sit in either one of the two specific sites. The commensurability of the molecular superlattice has also been established by high-resolution electron diffraction, according to which the herringbone lattice of Fig. 1a is compressed by about 2% from its bulk structure¹³.

It has been a long-standing question why a large molecule like PTCDA, covering more than 15 silver atoms, recognizes and locks

into a specific adsorption site on a surface which offers no preformed template structure. This puzzle has been compounded by the realization that the bonding is not accomplished locally by the oxygen atoms, but rather by the delocalized π -electron system²⁵. Yet perylene (see below), which has an identical π -electron backbone to that of PTCDA, does not show this epitaxy on Ag(111). Although it is clear that quadrupolar interactions between neighbouring molecules (as present in PTCDA, but not in perylene) can provide internal stabilization of the molecular layer, it is not at all obvious why this stabilized layer should lock into a strained substrate registry.

Here we present a solution to the epitaxy puzzle. It is based on results from first-principles quantum chemical calculations and electron energy-loss spectroscopy, a vibrational spectroscopy most sensitive to infrared-active modes of surface adsorbed molecules. The crucial spectrum is shown in Fig. 2a. Four strong modes (marked in red) are observed, which can be traced back to intrinsically Raman-active modes of the free molecule^{26,27}. At the surface, these acquire infrared activity, because on excitation of these vibrations electric charge is pumped back and forth between the metal surface and the molecule. This pumping action is caused by atomic displacements in the molecule and the ensuing redistributions of electronic probability densities. The question arises as to why a particular set of 4 out of 19 molecular Raman modes is enhanced selectively. In the past, this question could not be answered, in spite of a quite detailed understanding of the mechanism by which the modes are enhanced so strongly^{26,27}. We will show below that this problem is very closely related to the epitaxy puzzle.

The observed structure in Fig. 1a indicates that the molecule–substrate interaction (site recognition) and the intermolecular interaction (herringbone pattern) do not cause mutual frustration. Reflecting on the possibility of reconciling the two, one is led to the preliminary hypothesis of a localized reaction site in the centre of the molecule. Indeed, in the limiting case of a point-like molecular reaction centre, very little influence of the molecule–substrate interaction on the molecular orientation would be expected, because in that case the molecule could rotate almost freely around its central anchoring point on the surface in order to lock finally into the preferred herringbone orientation.

The reasoning behind our experimental approach to verify the existence of this anticipated molecular reaction centre is as follows. Because of a static charge transfer from the metal to the molecule, the former LUMO (lowest unoccupied molecular orbital, compare Fig. 1b) of the free molecule is the orbital most closely connected to the metallic charge reservoir²⁶. Whenever the electron density in a localized part of this spatially extended orbital is modulated by a vibrational excitation, we may expect that the metal electrons compensate this redistribution (leading to the above mentioned pumping action), under the condition that the respective part of the molecule is actually coupled strongly to the substrate. Conversely, if we determine which subsections of the molecule suffer significant charge modulations specifically for those four modes that are found enhanced in our experiments (Fig. 2a), we can work out which part of the molecule is coupled to the metallic charge reservoir most strongly, or, in other words, which part of the molecule plays the predominant role in forming the bond to the substrate.

To carry out this analysis, we have analysed *ab initio* density functional calculations of vibrational modes²⁸. Following our initial reasoning that the central carbon ring may play a crucial role, we have calculated the modulation of its area when Raman modes of the free molecule are excited (circles in Fig. 2a). Indeed, of the six modes for which the breathing amplitude in the central carbon ring is particularly large (red and blue circles), four coincide almost perfectly with the observed vibrations, if the calculated red shift²⁶ of the two asymmetric peaks is taken into account (red bars in Fig. 2a). This confirms the notion of the central carbon ring as the crucial

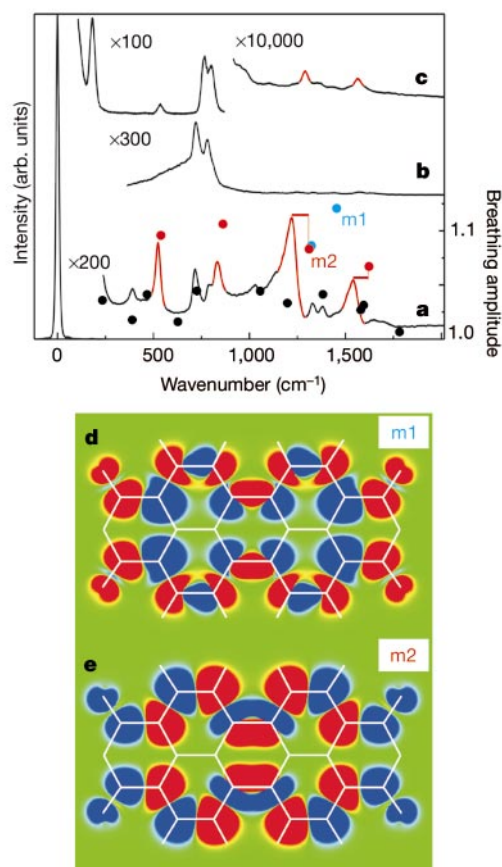


Figure 2 Spectroscopic signature of epitaxial site recognition. **a–c**, High-resolution electron energy-loss spectra for PTCDA and perylene on Ag(111). **a**, 0.3 monolayers of PTCDA. Peaks in red belong to the four Raman enhanced modes discussed in the text. Circles in **a** indicate the normalized breathing amplitude of the central carbon ring in PTCDA, as calculated from density functional theory of the free molecule²⁸. They are placed at the corresponding experimental Raman frequencies³⁰. Red circles, modes for which both the breathing amplitude is large and the charge modulation in the central ring is uncompensated. Blue circles, modes with large breathing amplitude but partially compensated charge modulation. Black circles, remaining modes. **b**, 0.3 monolayers of PTCDA deposited and measured at 100 K. Note that no Raman modes are enhanced. **c**, Perylene on Ag(111) also shows a slight activation of two Raman modes. Shown is a spectrum for a nominal coverage of a few monolayers. Additional data unambiguously prove that the red peaks derive from molecules in the first layer. **d, e**, Maps of vibration-induced charge density modulations in the LUMO of PTCDA²⁸ (arbitrary units). The maps are produced by (1) calculating the LUMO for the two turning points of the normal mode oscillation (Born–Oppenheimer approximation), (2) squaring it to get the charge distributions, and (3) subtracting the two turning point distributions from each other. Depicted is the cross-section 0.24 Å below the plane of the molecule. To reveal the general modulation pattern, the colour scale has been adjusted such that all areas in which the net change is larger than 1% of the maximum change are coloured homogeneously red (–) or blue (+). Maps **d** and **e** refer to the vibrational modes m1 and m2, as indicated.

reaction centre. If we turn to a map of vibration-induced charge density modulations in the LUMO (compare Fig. 2d, e), we can in fact establish a rigorous criterion for the correct prediction of four enhanced Raman modes (red circles): the central carbon ring must show both a strong breathing amplitude and an overall charge density modulation. For two examples, turn to Fig. 2: for the breathing mode in Fig. 2e, all parts within the central ring pump electrons synchronously, that is, with the same phase. On the other hand, the mode in Fig. 2d exhibits out-of-phase pumping for different parts of the central carbon ring, such that the overall density modulation is at least partially compensated within the molecule. In fact, we find that of the six breathing modes, the calculated modes marked by blue circles in Fig. 2 show partial compensation. This eliminates them from our list of modes for which strong enhancement is expected.

We have analysed all other Raman modes in the same way, and find that there is no other part of the molecule for which all four enhanced modes exhibit a mutually consistent charge modulation pattern. Thus, if other parts of the molecule were implicated in the bonding, a different subset of Raman modes would be enhanced in our spectroscopy data (note that other strong modes in Fig. 2a, apart from those marked in red, are infrared modes²⁶). In conjunction with the theoretical analysis, the spectrum in Fig. 2a therefore proves our hypothesis: the central carbon ring is the part of the molecule most effectively substrate-coupled. It constitutes the molecular reaction centre, which provides the large molecule with a lateral resolution high enough to single out and recognize a specific site on the close-packed silver surface to lock into. At the same time, the local nature of the reaction centre allows the molecular orientation to adjust to the intermolecular interactions, boosting the excellent stability of the PTCDA layers even further.

On the basis of the above site recognition model, we predict that epitaxy should not occur (1) if the reaction centre has been blocked, or (2) if its intrinsic activity is reduced. As we can monitor the reaction state of the centre spectroscopically, we can survey the correlation between interface structure and the recognition reaction. The outcome of these experiments is not only in complete agreement with our expectation, it also sheds additional light on the bonding mechanism and—most importantly—reveals a further technique to verify site recognition experimentally.

First, we consider PTCDA monolayers that were deposited on Ag(111) at 100 K. In this case, our STM results (not shown) indicate the complete absence of molecular order in general and of interface registry in particular. Coincidentally, in electron energy-loss spectra no enhanced Raman modes are observed (Fig. 2b). We therefore must conclude that a strong chemisorptive bond to the surface has not been formed, either by the central molecular reaction centre or by any other part of the molecule. Instead, the site recognition reaction is blocked by a lack of activation energy and the molecule is caught in a metastable chemisorption precursor. In this state, PTCDA ‘floats’ on the metallic electron density instead of reacting with it.

Second, we turn to a comparison between PTCDA and its backbone molecule perylene (see Fig. 3b inset) in order to investigate the influence of the side groups on the central reaction centre’s activity. For perylene monolayers on Ag(111), electron diffraction suggests an orientational ‘liquid’, in which the molecules are positionally ordered in an incommensurate close-packed superlattice but orientationally disordered and mobile. Having thus no indication of site recognition, we must ask whether the central reaction centre exists at all in bare perylene. The answer is again provided by electron energy-loss spectroscopy. We observe the same activated Raman peaks as for PTCDA (Fig. 2c). They are, however, orders of magnitude weaker than for PTCDA, indicating that, while a molecular reaction centre may still exist in the perylene backbone, its residual activity would be too small for the molecule to recognize a preferred site.

In the absence of direct (that is, structural) evidence for site recognition, can we be sure that the peaks in Fig. 2c really do indicate the existence of a reaction centre? To resolve this question, we must search for conditions under which the hypothetical reaction centre in perylene allows successful site recognition. These conditions can be found in an STM experiment. At first glance, the image of Fig. 3, recorded on perylene/Ag(111) near monolayer coverage, shows no features related to the molecular surface structure. We observe the lateral lattice periodicity of bare silver instead. Surprisingly, however, its corrugation is ~ 10 times larger than typically observed on Ag(111) (compare Fig. 3b and c).

The appearance of anomalously large corrugation in STM, in conjunction with several other features also observable in Fig. 3a—such as the fuzziness of the images in the fast scanning direction—is a clear indication that the imaging process between tip and surface is being influenced by an adsorbate²⁹. Careful experimentation in the present case can exclude any other adsorbate but perylene itself. Detailed analysis reveals that here we encounter a regime of molecularly sensitized tunnelling microscopy. Owing to weak molecule–substrate interaction, the tip gains control over a molecule and drags it across the surface, while the molecule (still) couples electronically with the substrate surface. The image contrast in the STM image of Fig. 3 is due to a variation in the mixing between molecular states and substrate states as the lateral position of the tip/molecule complex above the substrate surface changes. In

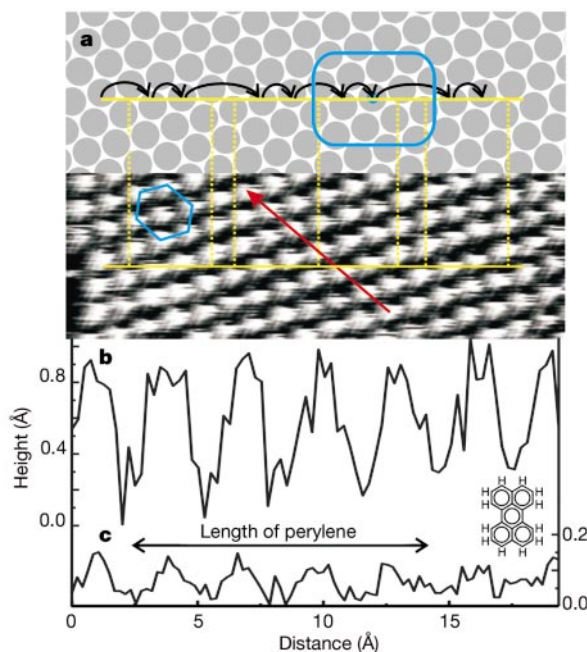


Figure 3 Probing site recognition with the STM. **a**, Bottom part: STM micrograph from the perylene/Ag(111) interface near monolayer coverage: $2 \times 5 \text{ nm}^2$, $I_t = 560 \text{ pA}$, $U_s = -0.15 \text{ V}$ (300 M Ω tunnelling resistance). Fast scanning direction is from left to right. In the vertical, the image has been scaled by a factor 0.86 in order to yield a nearly regular hexagonal shape of the unit cell (drift compensation). Top part: underlying atomic structure of the Ag(111) surface (phase relation to the bottom part is arbitrarily chosen). A scan line is shown as a solid yellow line, both in the STM micrograph and in the Ag(111) surface structure. Dotted yellow lines indicate tip positions with low local density of states. In the present plot, they correspond to positions near bridge sites. Molecular ‘jumps’ of perylene (blue outline drawn approximately to scale) are indicated by curved black arrows. For additional explanations, refer to the text. **b**, Line profile (height in Å) along the red arrow in **a**. **c**, Typical line profile measured on clean Ag(111) (corresponding image is not shown). Inset, structural formula of perylene.

fact, the observed strong variation of the electronic interaction between molecule and substrate may be interpreted as repeated bond formation and breaking. If we analyse a particular scan line (compare top and bottom parts of Fig. 3a), we find that a given local density of states (that is, brightness) is without exception associated with (nearly) equivalent lattice positions on the underlying substrate. Thus, the molecule 'jumps' from adsorption site to adsorption site, always in the direction defined by the motion of the tip, while the range of the jump is determined by the relative orientation of the scanning direction and the crystallographic axes. The tip never loses control over the molecule, but the latter transiently binds to the substrate before being moved on. The fact that in this way the underlying lattice of silver is imaged unambiguously signals recognition of a specific site by perylene on Ag(111) and is thus equivalent to commensurability in condensed layers. We have therefore found conditions under which the weak reaction centre of perylene allows site recognition. The spectroscopic data indicating a weak bond between Ag(111) and perylene are thus fully confirmed by an independent technique²⁹ which is able to probe site recognition reactions even in cases where these are in strong competition with other influences, for example, intermolecular interactions. In fact, the tip suppresses the latter by forcing the apex molecule to sample the surface potential systematically.

The comparison of PTCDA with perylene proves the feasibility of effective control over epitaxial site recognition. In both molecules, the perylene backbone merely supplies potential for an efficient recognition centre. This potential is fully developed only if electro-negative side groups, which increase the activity of the centre, are attached to the backbone. Side-group substitution is thus an important strategy for establishing control over organic epitaxy. In the present case, in going from perylene to PTCDA, the activity is enhanced by a factor of 300 if measured in terms of relative spectroscopic activities. We anticipate that this demonstrated insight into and control of site recognition sensitivities could be used not only in the context of organic epitaxy, but also whenever molecule-substrate interactions are intended as driving forces for structuring assemblies of large π -conjugated molecules at the nanoscale. □

Received 20 January; accepted 10 July 2003; doi:10.1038/nature01901.

1. Dodabalapur, A. Organic light emitting diodes. *Solid State Commun.* **102**, 259–267 (1997).
2. Horowitz, G. Organic field-effect transistors. *Adv. Mater.* **10**, 365–377 (1998).
3. Dimitrakopoulos, C. D. & Malenfant, P. R. L. Organic thin film transistors for large area electronics. *Adv. Mater.* **14**, 99–117 (2002).
4. Katz, H. E. *et al.* A soluble and air-stable organic semiconductor with high electron mobility. *Nature* **404**, 478–481 (2000).
5. Schoonfeld, W. A. *et al.* Coloumb-blockade transport in single-crystal organic thin-film transistors. *Nature* **404**, 977–980 (2000).
6. Crone, B. *et al.* Large-scale complementary integrated circuits based on organic transistors. *Nature* **403**, 521–523 (2000).
7. Dimitrakopoulos, C. D. *et al.* Low voltage organic transistors on plastic comprising high-dielectric constant gate insulators. *Science* **283**, 822–824 (1999).
8. Horowitz, G. *et al.* Role of semiconductor/insulator interface in the characteristics of π -conjugated-oligomer-based thin film transistors. *Synth. Metals* **51**, 419–424 (1992).
9. Hill, I. G. *et al.* Organic semiconductor interfaces: electronic structure and transport properties. *Appl. Surf. Sci.* **166**, 354–362 (2000).
10. Meyer zu Heringdorf, F. J. *et al.* Growth dynamics of pentacene films. *Nature* **412** (2001), 517–520.
11. Forrest, S. R. Ultrathin organic films grown by molecular beam deposition and related techniques. *Chem. Rev.* **97**, 1793–1896 (1997).
12. Hillier, A. C. & Ward, M. D. Epitaxial interactions between molecular overlayers and ordered substrates. *Phys. Rev. B* **54**, 14037–14051 (1996).
13. Umbach, E. *et al.* Surface architecture with large organic molecules: Interface order and epitaxy. *Surf. Sci.* **402–404**, 20–31 (1998).
14. Chen, Q. *et al.* Epitaxy of a crystalline organic semiconductor: Perylene/Cu(110). *Chem. Mater.* **14**, 743–749 (2002).
15. Fink, R. *et al.* Substrate-dependent lateral order in naphthalene-tetracarboxylic-dianhydride monolayers. *Phys. Rev. B* **60**, 2818–2826 (1999).
16. Soukopp, A. *et al.* Superstructure formation of large organic adsorbates on a metal surface: A systematic approach using oligothiophenes on Ag(111). *Phys. Rev. B* **58**, 13882–13894 (1998).
17. Tada, H. *et al.* Structural analysis of lead phthalocyanine ultrathin films grown on cleaved faces of alkali halides by reflection high energy electron diffraction. *Surf. Sci.* **268**, 387–396 (1992).
18. Zhang, Y. & Forrest, S. R. Mechanisms of quasicrystalline ordering at organic molecular thin film interfaces. *Phys. Rev. Lett.* **71**, 2765–2768 (1993).

19. Forrest, S. R. *et al.* Ultrahigh-vacuum quasicrystalline growth of model van der Waals thin films II. Experiment. *Phys. Rev. B* **49**, 11309–11321 (1994).
20. Hirose, Y. *et al.* Ordered, quasi-epitaxial growth of organic thin film on Se-passivated GaAs(100). *Appl. Phys. Lett.* **66**, 944–946 (1995).
21. Schmitz-Huebsch, T. *et al.* Epitaxial growth of 3,4,9,10-perylene-tetracarboxylic dianhydride on Au(111): A STM and RHEED study. *Phys. Rev. B* **55**, 7972–7976 (1997).
22. Hirose, Y. *et al.* Quasicrystalline growth of the organic molecular semiconductor 3,4,9,10-perylene-tetracarboxylic dianhydride. *Phys. Rev. B* **52**, 14040–14047 (1995).
23. Forrest, S. R. *et al.* Organic-on-organic semiconductor contact barrier diodes. II. Dependence on organic film and metal contact properties. *J. Appl. Phys.* **56**, 543–551 (1984).
24. Möbus, M. *et al.* Structure of perylene-tetracarboxylic-dianhydride thin films on alkali halide crystal substrates. *J. Cryst. Growth* **116**, 495–504 (1992).
25. Taborski, J. *et al.* NEXAFS investigations on ordered adsorbate layers of large aromatic molecules. *J. Electr. Spec. Rel. Phenom.* **75**, 129–147 (1995).
26. Tautz, F. S. *et al.* Strong electron-phonon-coupling at a metal/organic interface: PTCDA/Ag(111). *Phys. Rev. B* **65**, 125405 (2002).
27. Tautz, F. S. *et al.* A comparison of the chemisorption behaviour of PTCDA on different Ag surfaces. *Surf. Sci.* **502–503**, 176–184 (2002).
28. Frisch, M. *et al.* *Gaussian 98*, Rev. A6, (1998) (Gaussian, Inc., Pittsburgh).
29. Böhringer, M. *et al.* Scanning tunneling microscope-induced molecular motion and its effect on the image formation. *Surf. Sci.* **408**, 72–85 (1998).
30. Wagner, V. Raman analysis of ordered organic monolayers on metal surfaces. *Phys. Status Solidi A* **188**, 1297–1305 (2001).

Acknowledgements We thank V. Shklover and E. Umbach of the University of Würzburg for making output data of the quantum chemical calculations available to us, and for discussions. The work was financially supported by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.S.T. (s.tautz@iu-bremen.de).

Growth of early continental crust by partial melting of eclogite

Robert P. Rapp^{1*}, Nobumichi Shimizu² & Marc D. Norman³

¹Mineral Physics Institute and Department of Geosciences, State University of New York at Stony Brook, Stony Brook, New York 11794-2100, USA

²Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

³Research School of Earth Sciences, Australian National University, Canberra, ACT 0200, Australia

* Present address: Geodynamics Research Center, Ehime University, 2-5 Bunkyo-cho, Matsuyama 790-8577 Japan

The tectonic setting in which the first continental crust formed, and the extent to which modern processes of arc magmatism at convergent plate margins were operative on the early Earth, are matters of debate^{1,2}. Geochemical studies have shown that felsic rocks in both Archaean high-grade metamorphic ('grey gneiss') and low-grade granite-greenstone terranes are comprised dominantly of sodium-rich granitoids of the tonalite-trondhjemite-granodiorite (TTG) suite of rocks^{3–7}. Here we present direct experimental evidence showing that partial melting of hydrous basalt in the eclogite facies produces granitoid liquids with major- and trace-element compositions equivalent to Archaean TTG, including the low Nb/Ta and high Zr/Sm ratios of 'average' Archaean TTG⁸, but from a source with initially subchondritic Nb/Ta. In modern environments, basalts with low Nb/Ta form by partial melting of subduction-modified depleted mantle^{9,10}, notably in intraoceanic arc settings in the forearc^{11,12} and back-arc^{13,14} regimes. These observations suggest that TTG magmatism may have taken place beneath granite-greenstone complexes developing along Archaean intraoceanic island arcs by imbricate thrust-stacking¹⁵ and tectonic accretion¹⁶ of a diversity of

subduction-related terranes. Partial melting accompanying dehydration of these generally basaltic source materials at the base of thickened, 'arc-like' crust would produce compositionally appropriate TTG granitoids in equilibrium with eclogite residues.

The trace-element characteristics of Archaean granitoids are almost universally used to draw inferences on the origins of Earth's earliest continental crust^{1–7}. It is well established from melting/phase equilibria experiments on natural and synthetic basalts that liquids compositionally similar to TTG in terms of major-element oxide components form by low-moderate degree melting of mafic crust (basalt) in the amphibolite, garnet-amphibolite, or eclogite facies^{17–20}. These experiments have provided data on the identity and proportions of coexisting residual (crystalline) phases, allowing for trace-element modelling of TTG petrogenesis when combined with available experimentally determined mineral-melt partition coefficients (*D* values) for amphibole, clinopyroxene and garnet.

Recently, a number of experimental studies have reported mineral-melt *D* values specific to TTG petrogenesis, that is, for tonalitic liquids in equilibrium with amphibolitic or eclogitic phase assemblages^{21–23}, measured in synthetic or natural systems but generally at trace-element concentrations well above natural abundance levels, to facilitate analysis by a variety of microbeam analytical techniques. In particular, partitioning data for the elements Nb, Ta, Zr and Sm have been used to develop a model for the origin of Earth's earliest continental crust, in which it is argued that only low-degree (~1–15%) melting of recycled (that is, subducted) oceanic crust in the form of amphibolite, a relatively low-pressure equivalent of hydrous basalt, can account for the high Zr/Sm and low Nb/Ta ratios that are assumed to be representative of 'average' Archaean TTG⁸. These conclusions are predicated on the presumption of an 'average' basaltic source material for Archaean TTG with a chondritic Nb/Ta ratio, discounting alternative scenarios for the petrogenesis of TTG in which melting accompanying the progressive dehydration of overthickened and/or subducted mafic crust would peak at somewhat greater depths, where crystalline residues would be comprised of eclogite.

Here we report combined ion-microprobe and laser-ablation data from inductively coupled plasma mass spectrometry (ICP-MS) for a number of key trace elements (including Nb, Ta, Zr and Sm), measured at natural abundance levels, for TTG liquids formed by partial melting of natural metabasalt at 2–4 GPa, and coexisting with eclogitic residues (see Methods section). We argue that TTG melts in equilibrium with eclogite residues, while satisfying 'first-order' trace-element constraints, are better able to account for the major-element compositional features of TTG than are amphibolitic residues. Furthermore, if one allows for compositional variability in the putative basaltic source for TTG, as observed for basalts from a diversity of subduction-related environments, then TTG melts in equilibrium with eclogitic residues also pass the 'critical test'⁸, possessing appropriately low (yet variable) Nb/Ta ratios.

First-order geochemical characteristics of Archaean TTG include strong fractionation of light rare-earth elements (such as La) from heavy rare-earth elements (such as Yb), and Sr from Y, coupled to relative depletions in Yb and Y, resulting in high La/Yb and Sr/Y ratios^{1–7,17}. These are features shared by the experimental TTG melts in equilibrium with eclogite, as well as their mantle-hybridized counterparts (Fig. 1), and are explicable in terms of the *D* values for these elements between eclogitic phases and TTG melts. For example, Sr and La are highly incompatible in both garnet and clinopyroxene (that is, mineral-melt *D* values $\ll$ 1.0), whereas the heavy rare-earth elements (for example, Dy, Er and Yb) and Y are strongly compatible in garnet (*D* values $\gg$ 1.0). As a consequence, TTG granitoids in equilibrium with eclogite possess characteristically high La/Yb and Sr/Y ratios^{1,3–7,17}. By contrast, the *D* values for

La and Sr between amphibole and TTG melt are an order of magnitude higher than those between garnet and TTG melt, and the *D* values for Y and Yb are an order of magnitude lower than those between garnet and TTG melt. Thus an amphibole-dominated residue will be less capable of imparting the high Sr/Y and La/Yb ratios and strong depletions in Yb and Y that are characteristic of Archaean TTG^{1,5–7}. Our experimental liquids fall well within the field defined by early-late Archaean TTG for other trace elements as well, for example, Th and U (Fig. 2).

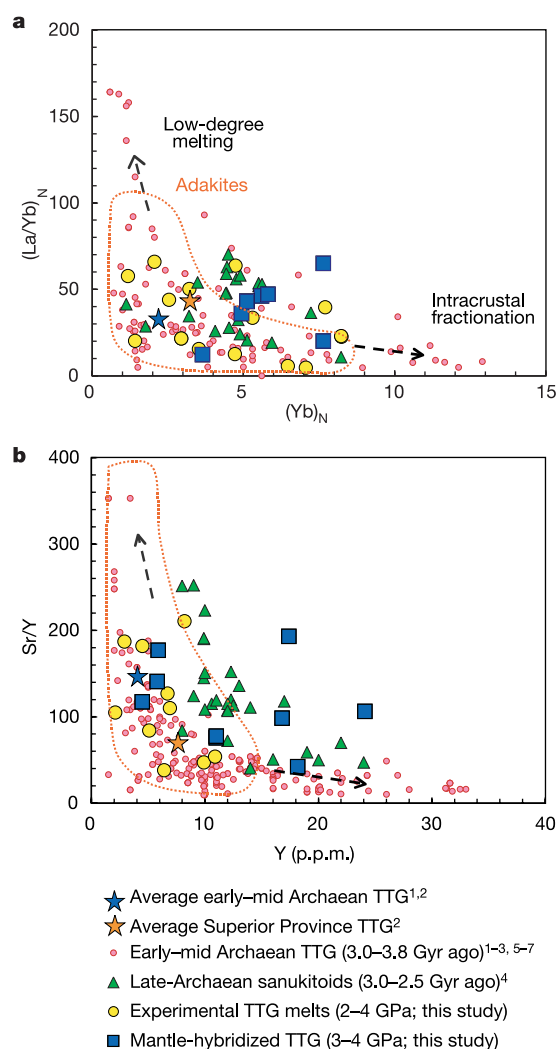


Figure 1 Selected trace-element ratios in early-mid-Archaean TTG and late Archaean sanukitoids^{1–7} match those of experimental liquids in equilibrium with eclogite. Data for 'pristine' TTG melts in equilibrium with eclogite (Mg-number $\approx$ 0.30–0.50) at 2–4 GPa (refs 18, 20), and 'mantle-hybridized' TTG (Mg-number $\approx$ 50–70) in equilibrium with garnet websterite mineral assemblages at 3.5–4.0 GPa from ref. 19 and R.P.R. and N.S., unpublished data. Also shown is the field for adakite lavas (orange dotted line) in modern subduction zones¹, and the predicted effects of intracrustal differentiation by fractionation of amphibole and plagioclase at low pressure (black arrows)^{1,5,6}. All experimental data for 'mantle-hybridized' from ref. 19 and R.P.R. and N.S., unpublished data; and for TTG liquids at 1.2–3.2 GPa from ref. 18, and unpublished data for additional peridotite assimilation experiments at 3–4 GPa. **a**, Plot of $(La/Yb)_N$ versus $(Yb)_N$, normalized to chondritic values. Extreme values of $(La/Yb)_N$ result from very low-degree melting of metabasaltic sources, and low $(La/Yb)_N$ ratios coupled with low Yb abundances from intracrustal fractionation of plagioclase and amphibole¹. **b**, Plot of Sr/Y versus Y for the same data sets. Bulk distribution coefficients for TTG melts in equilibrium with eclogite residues account for the highly incompatible behaviour of both La and Sr, and the highly compatible behaviour of Y and Yb, which are strongly partitioned into garnet^{8,21–23}.

The major-element compositions of the great majority of Archaean TTG are characterized by SiO_2 contents of 65–73 wt% and Mg-numbers of 0.30–0.50 (where Mg-number = molar $\text{Mg}/(\text{Mg} + \text{Fe}^{2+})$), and alumina saturation indexes (ASI, or A/CNK) of approximately 1.0 ± 0.05 (where A/CNK = molar $\text{Al}_2\text{O}_3/(\text{CaO} + \text{Na}_2\text{O} + \text{K}_2\text{O})^{1-7}$). These features (in particular, A/CNK ratios of ~ 1.0) require relatively high degrees of melting (>10 – 15%) of a garnet-bearing basaltic protolith, at pressure–temperature (P – T) conditions close to or beyond the ‘amphibole-out’ phase boundary²⁰, where amphibole will either be entirely absent from the eclogite residue, or a minor phase. By contrast, amphibole-dominated residues of hydrous basalt melting possess Mg-numbers generally below 0.30, and A/CNK ratios >1.1 (refs 17, 18, 20).

The argument for an origin for TTG by partial melting of an amphibolite source hinges largely on the ability of amphibole to fractionate Nb from Ta⁸, and thereby to explain the low (subchondritic) Nb/Ta ratio of ‘average’ Archaean TTG, by assuming a source with initially chondritic Nb/Ta. In fact, as argued below, Archaean TTG, as well as their potential basaltic source materials, show a very wide range in Nb/Ta ratios.

A trend to lower Nb/Ta ratios has been documented for near-axis seamounts in the eastern Pacific⁹, for oceanic crust formed in back-arc basins^{13,14}, for some boninites formed in the fore-arc region^{11,12}, and for ‘boninite-like’ basalts in Archaean greenstone belts^{24,31}, including 3.7–3.8 Gyr old mafic volcanics from the Isua greenstone belt in West Greenland (ref. 31 and A. Polat, personal communication, 2002). In the Eastern Pacific seamounts, this trend is attributed to contamination of the upper mantle source region for these basalts by material that had been ‘processed’ through a subduction zone⁹, and for back-arc basin basalts, boninites and boninite-like rocks in Archaean greenstone belts, to enrichment of their mantle sources by slab-derived fluids possessing low Nb/Ta

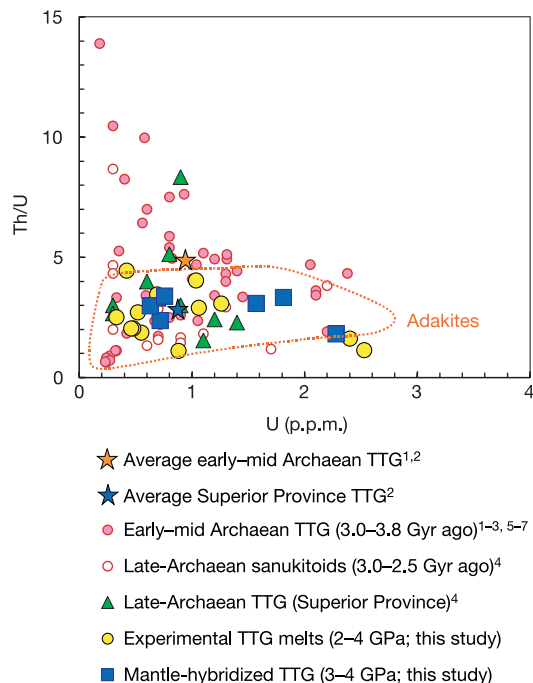


Figure 2 Variations of Th/U versus U in Archaean TTG are also consistent with their equilibration with eclogitic residues. Data are shown for both ‘pristine’ and ‘mantle-hybridized’ TTG melts formed at 2–4 and 3.5–4.0 GPa, respectively, measured by ion microprobe and laser-ablation ICP-MS, and for early–mid-Archaean TTG, late-Archaean TTG and sanukitoids (Mg-rich monzodiorites⁴) from the Slave and Superior provinces of Canada.

(refs 9–12). We consider our starting basalt AB-1 (see Table 1 and ref. 19) to be compositionally representative of basalts formed in marginal (back-arc) basins, generally subduction-enriched oceanic crust, boninitic lavas from fore-arc environments, and some greenstone belt metavolcanics that resemble these rocks compositionally

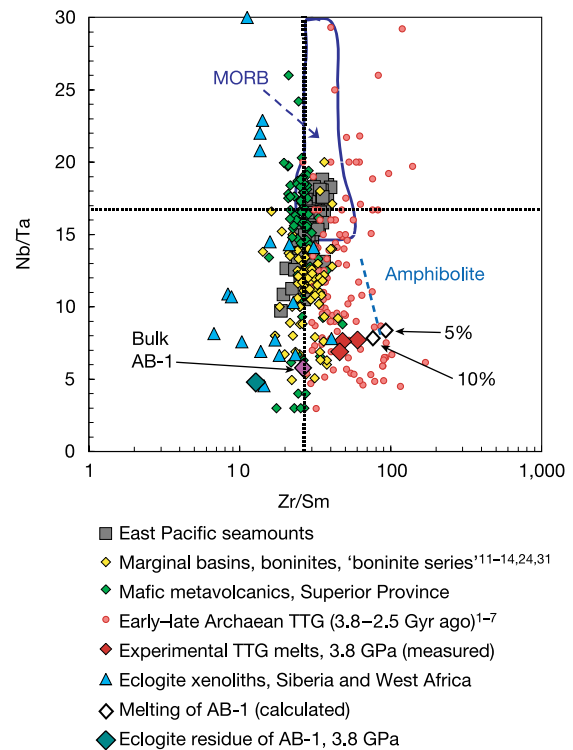


Figure 3 Niobium/tantalum versus zirconium/samarium ratios suggest a compositionally variable oceanic basaltic source for early–late-Archaean TTG magmas. Archaean TTGs include data from West Greenland (3.8 Gyr ago), the Kaapvaal (3.5–3.6 Gyr ago) and Zimbabwean (3.4 Gyr ago) cratons, the West African Craton (~ 3.5 Gyr ago), the Pilbara Craton (~ 3.5 Gyr ago), and the late-Archaean (~ 2.5 – 3.0 Gyr ago) Slave and Superior provinces of Canada^{1–7}. These data are shown relative to possible basaltic source materials in Archaean greenstone belts, and their modern analogues, including seamounts from the eastern Pacific⁹, intraoceanic basalts, including marginal basin basalts from the western Pacific^{13,14}, Phanerozoic boninites^{11,12} and Archaean boninite-like or boninite series rocks^{24,31}. Experimental data are shown for experimental TTG liquid formed by 30% melting of AB-1 and mantle-hybridized equivalents of the TTG-liquid at 30% melting, formed by assimilation of peridotite¹⁹ (measured by laser ablation ICP-MS; see Table 1), as well as for 10% and 5% melting of AB-1 (calculated). Range of ratios from partitioning calculations of ref. 8 for 1–15% partial melting of amphibolite at 0.8 GPa are shown as light-blue bars. Initial variations in the Nb/Ta ratio of parental TTG magmas would reflect primarily the (variable) Nb/Ta ratio of their basaltic source(s). For TTG derived by eclogite melting of basaltic sources with bulk Nb/Ta ratios higher than AB-1 (~ 7 – 17), post-emplacement fractionation of amphibole would be able to reduce the Nb/Ta ratio of derivative magmas somewhat. Alternatively, small amounts of residual amphibole in the source (that is, melting of garnet amphibolite) would produce TTG liquids with Nb/Ta ratios slightly lower than the initial Nb/Ta ratio of the source. Eclogites interpreted as residues of TTG magmatism^{26,27} show a range of Nb/Ta ratios consistent with our model, in which the basaltic source protolith possesses a range of initial Nb/Ta ratios. This protolith could be mafic arc crust formed in intraoceanic subduction zones, or oceanic crust whose source itself contains components of crust that had been recycled, perhaps repeatedly, through a more primitive subduction regime in the early Archaean or Hadean. The amphibolite melting model of ref. 8 would predict a clustering of eclogite residues with superchondritic Nb/Ta ratios, which is not observed; instead, eclogites interpreted as residues of TTG magmatism display a wide range of Nb/Ta ratios, reflecting the variability of their basaltic (that is, oceanic) crustal source. Thus the complementarity between TTG granitoids and eclogite residues is primarily ‘east-to-west’, and not ‘north-to-south’, as the amphibolite melting model would predict. The field for MORB, mid-ocean ridge basalt, is bounded by the blue curve.

Table 1 **Abundances of trace elements in starting material and TTG melts**

Sample	Nb	Ta	Nb/Ta	Zr	Sm	Zr/Sm
Bulk AB-1	3.2 ± 0.1	0.51 ± 0.01	6.3	80 ± 2	2.99 ± .01	26.7
Melt of AB-1 (<i>n</i> = 3)	5.4 ± 0.6	0.7 ± 0.2	7.7	272 ± 15	4.5 ± 0.5	60
Melt of AB-1 + 15%DP (<i>n</i> = 3)	9.7 ± 0.1	1.4 ± 0.1	6.9	320 ± 30	7.7 ± 1.1	42
Melt of AB-1 + 12%FP (<i>n</i> = 3)	11.4 ± 0.3	1.5 ± 0.1	7.6	358 ± 10	7.3 ± 0.2	49

DP, depleted peridotite; FP, fertile peridotite. Data given for pristine and mantle-hybridized TTG melts of basalt (AB-1) at 3.8 GPa.

(Fig. 3). The actual process causing fractionation of Nb from Ta and leading to subchondritic Nb/Ta ratios remains a contentious issue^{9,10,13}. One possible explanation involves subduction-related metasomatism of depleted oceanic lithosphere of the mantle wedge by slab-derived fluids possessing low Nb/Ta, acquired by equilibration with rutile-bearing oceanic crust^{9,25}. In this scenario, the partition coefficient for Nb between rutile and aqueous fluids is higher than that for Ta, imparting a low Nb/Ta ratio to the fluid. Subsequent melting of this fluid-metasomatized, depleted oceanic mantle lithosphere would produce basalts with variable but generally subchondritic Nb/Ta ratios, representing possible mafic sources for TTG magmas (Fig. 3).

Low- to moderate-degree (5–30%) melting of basalt AB-1 at 3–4 GPa produces TTG liquids with low Nb/Ta ratios and elevated Zr/Sm ratios (the latter being due to higher partition coefficients for Sm in clinopyroxene relative to Zr), in equilibrium with rutile-bearing eclogite residues. The variable but generally low Nb/Ta and high Zr/Sm ratios of Archaean TTG can be explained by partial melting of oceanic basalt with variable but generally sub-chondritic Nb/Ta ratios, leaving residues consisting of eclogite which also possess low but variable Nb/Ta (Fig. 3). The presence of residual rutile raises the Nb/Ta ratio of the partial melt of AB-1 only slightly, due to the somewhat higher *D* for Ta relative to Nb in rutile in the presence of granulite melts²⁵. From mass balance considerations, we estimate that rutile-TTG melt *D* values are ~55–60 for Nb, and 85–90 for Ta, with a $D^{\text{Nb/Ta}} \approx 0.68$, comparable to previous estimates²⁵ of $D^{\text{rutile/melt}} \approx 27$ for Nb and $D^{\text{rutile/melt}} \approx 44$ for Ta, with a $D^{\text{Nb/Ta}} \approx 0.60$ –0.67. It appears that the $D^{\text{rutile/melt}}$ used in ref. 8 in modelling melting of rutile-bearing eclogite were much higher, especially for Nb; combined with a starting material with an assumed chondritic Nb/Ta ratio, their calculations suggested that partial melting of rutile-bearing eclogite would result in TTG liquids with superchondritic Nb/Ta. Note, however, that the subchondritic Nb/Ta ratios of our experimental eclogite residues are comparable to those of rutile-bearing eclogites previously interpreted as being residues from TTG magma generation^{26,27}. The generally subchondritic (but highly variable) Nb/Ta ratios of Archaean TTG, consequently, require neither an amphibolitic source⁸, nor a hidden eclogitic reservoir with superchondritic Nb/Ta (ref. 28), complementary to ‘average’ Archaean TTG (with subchondritic Nb/Ta) and linked to continent formation. We argue that the use of an ‘average’ Nb/Ta ratio for Archaean TTG is misleading, since the complementary nature of TTG and their eclogitic residues spans a broad range of Nb/Ta ratios, reflecting the compositional diversity of their basaltic source material.

Models for the origin and growth of the continental crust and the chemical evolution of the subcratonic lithosphere in the Archaean require a general understanding of the genetic relationship between TTG granitoids comprising the earliest continental masses, and their deep roots or ‘keels’ in the underlying mantle. Our results suggest that some eclogitic xenoliths and inclusions in diamonds from the subcratonic mantle may indeed represent residues from partial melting and TTG magmatism^{26,27,29}. The source for Archaean TTG magmas may have been oceanic crust with variable Nb/Ta ratios, formed in a variety of environments in association with intraoceanic subduction; tectonic imbrication and accretion of

these disparate terranes would have led to partial melting and TTG magmatism at the base of overthickened, juvenile arc crust^{15,16}. We see TTG magmas, then, as the result of a process that begins with tectonic thickening of oceanic crust in a subduction-style regime, followed by progressive metamorphism and partial melting of metabasaltic sources, initially composed of amphibolite and garnet amphibolite, but culminating in TTG liquids that for the most part were in equilibrium with anhydrous eclogite residues, before mobilization and emplacement in evolving granite-greenstone complexes that would ultimately form the first continents. □

Methods

Melting experiments on hydrous basalts produce TTG liquids over a broad pressure interval (1–4 GPa), corresponding to the breakdown of water-bearing minerals (primarily amphibole, but also zoisite, lawsonite and phengite). Our own experiments were conducted on natural amphibolitic rock powders in the piston-cylinder apparatus ($P = 1.8$ –3.2 GPa; refs 18, 20) or the multi-anvil apparatus ($P = 3.5$ –3.8 GPa; ref. 19), at temperatures and pressures close to or beyond the amphibole-out phase boundary²⁰. At the P – T conditions of the experiments, partial melting produces 10–30% (by weight) TTG liquid coexisting with eclogite (garnet + Na-rich clinopyroxene) residues containing minor amounts (~0.5–1.0 wt%) of rutile, and with amphibole either absent, or present as a minor phase in the lower-pressure experiments^{18,20}. In related experiments in the multi-anvil apparatus at 3.8 GPa, ‘pristine’ TTG liquids produced by partial melting of metabasalt at 1,100 °C were allowed to infiltrate and react with an overlying layer of natural peridotite, producing ‘mantle-hybridized’ TTG liquids in equilibrium with garnet websterite reaction residues (garnet + orthopyroxene ± clinopyroxene)¹⁹.

Selected trace elements were measured in the granitoid liquids produced in these experiments, using primarily the ion microprobe, with a number of samples also analysed by laser-ablation ICP-MS. From limited ion probe analyses of coexisting crystalline phases, mineral-melt partition coefficients (*D* values) between TTG melts and eclogitic (garnet and clinopyroxene) residues were also determined (R.P.R. and N.S., manuscript in preparation). Mineral-melt *D* values between garnet and TTG liquids, and between clinopyroxene and TTG liquids, measured at natural abundance levels, are similar to those previously published for ‘doped’ experiments for a broad range of trace elements^{21–23}. These measurements not only provide an independent test of Henry’s law behaviour, but more importantly, they indicate that the abundances of trace elements measured in the TTG liquids formed in rock-melting experiments represent equilibrium concentrations, imposed by the crystal-chemical partitioning effects of the dominant phases (garnet and clinopyroxene).

One of the metabasalt compositions used in our experiments (AB-1) comes from extrusive sequences of the Josephine ophiolite complex of northern California and Oregon, which is interpreted as having formed by episodic spreading in a Late Jurassic back-arc basin. This particular sample has an initial Nb/Ta ratio of ~6.3 (Nb = 3.2 ± 0.1 p.p.m.; Ta = 0.51 ± 0.01 p.p.m.), and an initial Zr/Sm ratio of ~27 (Zr = 80 ± 2 p.p.m.; Sm = 2.99 ± 0.01 p.p.m.), as determined by solution ICP-MS (Table 1). Although we suspect that some contamination with respect to Ta may have taken place years ago during initial crushing of the sample in tungsten carbide (G. Harper, personal communication, 2002), artificially lowering the Nb/Ta ratio, this has no effect over the partitioning of these elements between solid and liquid phases during partial melting (essentially representing a slight doping of the basaltic source with Ta, at near-natural abundance levels).

With regard to estimates of mineral-melt *D* values between rutile and TTG liquids for Ta and Nb, we have shown that melts in equilibrium with amphibolite-dominated crystalline residues possess higher A/CNK ratios than those in equilibrium with eclogitic residues (the tonalite in the experiments of ref. 8 has a peraluminous A/CNK of ~1.15, compared to metaluminous A/CNK ratios of ~1.0 in our experiments. We believe that our estimated rutile-melt *D* values for Nb and Ta are much more reasonable values, because *D* values for Nb and Ta between rutile and granitoid melts have been shown to be strongly dependent on melt composition, and decrease dramatically as liquid composition changes from peraluminous to metaluminous³⁰.

Received 23 March; accepted 28 August 2003; doi:10.1038/nature02031.

1. Martin, H. Adakitic magmas: modern analogues of Archean granitoids. *Lithos* **46**, 411–429 (1999).
2. Smithies, R. H. The Archean tonalite-trondhjemite-granodiorite (TTG) series is not an analogue of Cenozoic adakite. *Earth Planet. Sci. Lett.* **182**, 115–125 (2001).

3. Hunter, D. R., Smith, D. G. & Sleight, D. W. W. Geochemical studies of Archean granitoid rocks in the Southeastern Kaapvaal Province: implications for crustal development. *J. Afr. Earth Sci.* **15**, 127–151 (1992).
4. Feng, R. & Kerrich, R. Geochemical evolution of granitoids from the Archean Abitibi southern volcanic zone and the Pontiac subprovince, Superior Province, Canada: implications for the tectonic history and source regions. *Chem. Geol.* **98**, 23–70 (1992).
5. Luais, B. & Hawkesworth, C. J. The generation of continental crust: an integrated study of crust-forming processes in the Archean of Zimbabwe. *J. Petrol.* **35**, 43–93 (1994).
6. Nutman, A. P., Bennett, V. C., Friend, C. R. L. & Norman, M. D. Meta-igneous (non-gneissic) tonalites and quartz-diorites from an extensive ca. 3800 Ma terrain south of the Isua supracrustal belt, southern West Greenland: constraints on early crust formation. *Contrib. Mineral. Petrol.* **137**, 364–388 (1999).
7. Kreissig, K., Nagler, T. F., Kramers, J. D., van Reenen, D. D. & Smit, C. A. An isotopic and geochemical study of the northern Kaapvaal Craton and the Southern Marginal Zone of the Limpopo Belt: are they juxtaposed terranes? *Lithos* **50**, 1–25 (2000).
8. Foley, S., Tiepolo, M. & Vannucci, R. Growth of early continental crust controlled by melting of amphibolite in subduction zones. *Nature* **417**, 637–640 (2002).
9. Niu, Y. & Batiza, R. Trace element evidence from seamounts for recycled oceanic crust in the Eastern Pacific mantle. *Earth Planet. Sci. Lett.* **148**, 471–483 (1997).
10. Weyer, S., Munker, C. & Mezger, K. Nb/Ta Zr/Hf and REEs in the depleted mantle: implications for the differentiation history of the crust-mantle system. *Earth Planet. Sci. Lett.* **205**, 309–324 (2003).
11. Bedard, J. H. Petrogenesis of boninites from the Betts Cove Ophiolite Newfoundland, Canada: identification of subducted source components. *J. Petrol.* **40**, 1853–1889 (1999).
12. Cluzel, D., Aitchison, J. C. & Picard, C. Tectonic accretion and underplating of mafic terranes in the Late Eocene intraoceanic forearc of New Caledonia (Southwest Pacific): geodynamic implications. *Tectonophysics* **340**, 23–59 (2001).
13. Munker, C. Nb/Ta fractionation in a Cambrian arc/back arc system. New Zealand: source constraints and application of refined ICPMS techniques. *Chem. Geol.* **144**, 23–45 (1998).
14. Nicholson, K. N., Picard, C. & Black, P. M. A comparative study of Late Cretaceous ophiolitic basalts from New Zealand and New Caledonia: implications for the tectonic evolution of the SW Pacific. *Tectonophysics* **327**, 157–171 (2000).
15. de Wit, M. J. On Archean granites, greenstones, cratons, and tectonics: does the evidence demand a verdict? *Precamb. Res.* **91**, 181–226 (1998).
16. Kusky, T. M. & Polat, A. Growth of granite-greenstone terranes at convergent margins, and stabilization of Archean cratons. *Tectonophysics* **305**, 43–73 (1999).
17. Sen, C. & Dunn, T. Dehydration melting of a basaltic composition amphibolite at 1.5 and 2.0 GPa: implications for the origin of adakites. *Contrib. Mineral. Petrol.* **117**, 394–409 (1994).
18. Rapp, R. P. & Watson, E. B. Dehydration melting of metabasalt at 8–32 kbar: implications for continental growth and crust-mantle recycling. *J. Petrol.* **36**, 891–931 (1995).
19. Rapp, R. P., Shimizu, N., Norman, M. D. & Applegate, G. S. Reaction between slab-derived melts and peridotite in the mantle wedge: experimental constraints at 3.8 GPa. *Chem. Geol.* **160**, 335–356 (1999).
20. Rapp, R. P. Amphibole-out phase boundary in partially melted metabasalt, its control over liquid fraction and composition, and source permeability. *J. Geophys. Res.* **100**, 15601–15610 (1995).
21. Barth, M. G., Foley, S. F. & Horn, I. Partial melting in Archean subduction zones: constraints from experimentally determined trace element partition coefficients between eclogitic minerals and tonalitic melts under upper mantle conditions. *Precamb. Res.* **113**, 323–340 (2002).
22. Klein, M., Stosch, H.-G., Seck, H. A. & Shimizu, N. Experimental partitioning of high field strength and rare earth elements between clinopyroxene and garnet in andesitic to tonalitic systems. *Geochim. Cosmochim. Acta* **64**, 99–115 (2000).
23. Klemme, S., Blundy, J. D. & Wood, B. J. Experimental constraints on major and trace element partitioning during partial melting of eclogite. *Geochim. Cosmochim. Acta* **66**, 3109–3123 (2002).
24. Boily, M. & Dion, C. Geochemistry of boninite-type volcanic rocks in the Protet-Evans greenstone belt, Opatica subprovince, Quebec: implications for the evolution of Archean greenstone belts. *Precamb. Res.* **115**, 349–371 (2002).
25. Green, T. H. Significance of Nb/Ta as an indicator of geochemical processes in the crust-mantle system. *Chem. Geol.* **120**, 347–359 (1995).
26. Jacob, D. E. & Foley, S. F. Evidence for Archean ocean crust with low high field strength element signature from diamondiferous eclogite xenoliths. *Lithos* **48**, 317–336 (1999).
27. Barth, M. G. *et al.* Geochemistry of xenolithic eclogites from West Africa. Part I: A link between low MgO eclogites and Archean crust formation. *Geochim. Cosmochim. Acta* **65**, 1499–1527 (2001).
28. Rudnick, R. L., Barth, M., Horn, I. & McDonough, W. F. Rutile-bearing refractory eclogites: missing link between continents and depleted mantle. *Science* **287**, 278–281 (2000).
29. Rollinson, H. Eclogite xenoliths in west African kimberlites as residues from Archean granitoid crust formation. *Nature* **389**, 173–176 (1997).
30. Horing, W.-S. & Hess, P. C. Partition coefficients of Nb and Ta between rutile and anhydrous haplogranite melts. *Contrib. Mineral. Petrol.* **138**, 176–185 (2000).
31. Polat, A., Hofmann, A. W. & Rosing, M. T. Boninite-like volcanic rocks in the 3.7–3.8 Ga Isua greenstone belt, West Greenland: geochemical evidence for intra-oceanic subduction zone processes in the early Earth. *Chem. Geol.* **184**, 231–254 (2002).

Acknowledgements This research was supported by National Science Foundation grants to R.P.R. and the Mineral Physics Institute at Stony Brook. We thank M. Defant for critical comments. Discussions with H. Martin, A. Polat, H. Rollinson, M. Rosing, H. Smithies and M. de Wit helped to improve the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.P.R. (rrapp@notes.cc.sunysb.edu).

Large Cretaceous sphenodontian from Patagonia provides insight into lepidosaur evolution in Gondwana

Sebastián Apesteguía & Fernando E. Novas

Museo Argentino de Ciencias Naturales “Bernardino Rivadavia”,
Av. Ángel Gallardo 470, Buenos Aires (1405), Argentina

Sphenodontian reptiles successfully radiated during Triassic and Jurassic times, but were driven almost to extinction during the Cretaceous period¹. The sparse Early Cretaceous record of sphenodontians has been interpreted as reflecting the decline of the group in favour of lizards, their suspected ecological successors². However, recent discoveries in Late Cretaceous beds in Patagonia partially modify this interpretation. Numerous skeletons of a new sphenodontian, *Príosphenodon avelasi* gen. et sp. nov., were collected from a single locality in the Cenomanian–Turonian Candeleros Formation, where it is more abundant than any other tetrapod group recorded in the quarry (for example, Crocodyliformes, Serpentes, Dinosauria and Mammalia)^{3,4}. Adult specimens of *Príosphenodon* reached one metre in length, larger than any previously known terrestrial sphenodontian. Here we propose, using available evidence, that sphenodontians were not a minor component of the Cretaceous terrestrial ecosystems of South America, and that their ecological replacement by squamates was delayed until the early Tertiary. The new discovery helps to bridge the considerable gap in the fossil record (around 120 million years) that separates the Early Cretaceous sphenodontians^{5–7} from their living relatives (*Sphenodon*).

Today, lepidosaurs are a diverse group of reptiles. With more than 6,000 species of lizard, snake, amphisbaenian and tuatara, they form an important part of the terrestrial vertebrate fauna. Although the evolutionary history of lepidosaurs during the late Mesozoic is relatively well known in North America and Asia, their record in the Southern Hemisphere (Gondwana) is very limited. As a result, basic questions about their evolutionary history remain incompletely answered, including the origin of snakes, the biogeography of iguanian lizards, the origin of basal squamates and the relictual presence of *Sphenodon*, the last sphenodontian, in New Zealand, an isolated part of Gondwana.

New findings presented below, including well-preserved and abundant skeletons of a new sphenodontian collected from Cenomanian–Turonian⁸ rocks in North Patagonia, permit a better understanding of some aspects of late Mesozoic lepidosaurian evolution in the southern continents.

Lepidosauria Haeckel, 1866 (ref. 9)

Sphenodontia Williston, 1925 (ref. 10)

Opisthodontia nov.

Eilenodontinae Rasmussen & Callison, 1981 (ref. 11)

Príosphenodon avelasi gen. et sp. nov.

Holotype. MPCA 300 (Museo Carlos Ameghino, Cipolletti, Río Negro Province, Argentina) consists of a partially articulated, adult skeleton (Fig. 1a–c).

Referred material. The present study was based on the holotype as well as the following specimens: MPCA 301, a juvenile incomplete skeleton (Fig. 1d, e); MPCA 302, postcranial isolated remains (Fig. 1g, h); and MPCA 303, a partially articulated adult skeleton. More than a hundred uncatalogued specimens, including isolated elements and complete skeletons, remain unprepared.

Locality and horizon. Upper layers of the Candeleros Formation (Cenomanian–Turonian)⁸, ‘La Buitrera’ fossil quarry, Cerro Policia, Río Negro Province, northwest Patagonia, Argentina. The specimens were found in fluvial sandstones, in association with snakes, turtles, araripesuchid crocodyliforms, theropod and saur-
 opod dinosaurs, mammals and dipnoan fishes^{4,12,13–15}.

Diagnosis. *Priosphenodon* differs from other sphenodontians in the following combination of features: robust skeleton with adult up to 1 m long; single, sharp, beak-like structure on the rostral end of the skull; bulging nasals; frontals rostrally straight; jugals dorsoventrally deep; suborbital fenestrae closed by transversely oriented ectopterygoids; maxillary and palatine tooth rows elongate, straight and in parallel; palatal roof narrow; teeth closely packed, with imbricate labial and lingual flanges; no caniniform teeth and no dental regionalization; and square,

distally expanded unguals.

Description. The skull of *Priosphenodon* (Fig. 1a–e) is triangular, deep and massive, ending rostrally in a single, sharply pointed beak-like structure that is presumably conformed by fused premaxillary teeth. Both pre- and postfrontals are profusely sculptured, a condition shared with the Late Jurassic *Eilenodon robustus* from North America (Los Angeles County Museum, LACM 120462). The parietal crest is narrow, as in derived sphenodontians (for example, *Sphenodon*, *Zapatadon*, *Kallimodon* and *Sapheosaurus*). The jugal is remarkably deep dorsoventrally and is devoid of the typical caudal bifurcation present in most sphenodontians¹⁶, thus conforming a tall supratemporal bar, as in *Homoiosaurus*¹⁷. This novel morphology of the temporal region provided an enlarged area for adductor muscle attachment.

The palatal roof of *Priosphenodon* is narrower than that of any

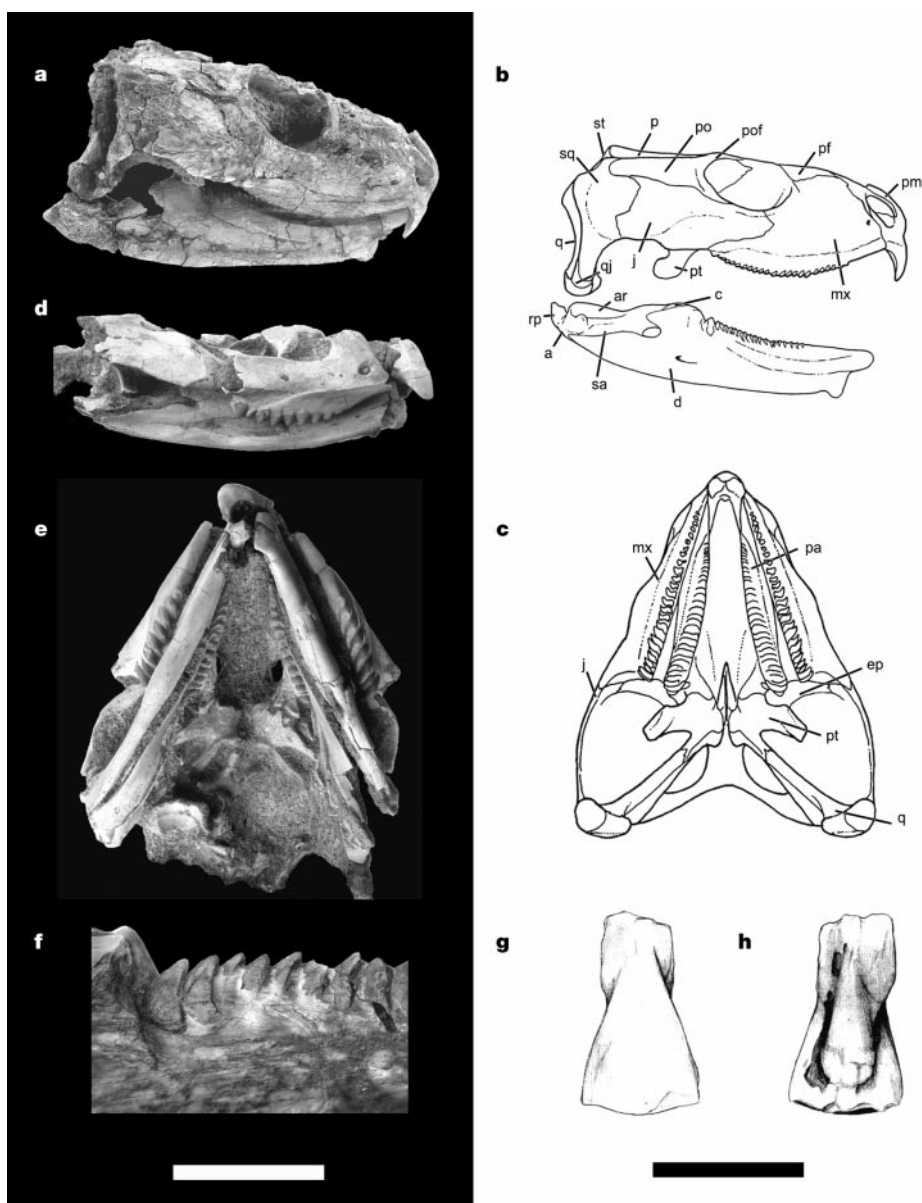


Figure 1 *Priosphenodon avelasi*. **a–c**, MPCA 300, holotype. **a**, Skull in lateral view. **b**, Skull and jaw in lateral view. **c**, Skull in ventral view. **d**, **e**, Skull of a juvenile specimen (MPCA 301) in lateral (**d**) and ventral (**e**) views. **f**, Detail of a sequence of teeth of the right jaw in lateral view. **g**, **h**, Ungual phalanges (MPCA 302) in dorsal (**g**) and ventral (**h**) views. Scale bar, 40 mm (**a–c**); 16 mm (**d**, **e**); 10 mm (**f–h**). a, angular; ar, articular; c, coronoid;

d, dentary; ep, ectopterygoid; j, jugal; mx, maxilla; p, parietal; pa, palatine; pf, prefrontal; pm, premaxilla; po, postorbital; pof, postfrontal; pt, pterygoid; q, quadrate; qj, quadratojugal; rp, retroarticular process; sa, surangular; sq, squamosal; st, supratemporal.

other terrestrial sphenodontian, and is framed by straight palatine tooth rows that run parallel to the maxillary rows. Expanded and transversely oriented ectopterygoids close the suborbital fenestrae, a feature unique among Sphenodontia. The lower jaw is deep, with low coronoid and short retroarticular processes, as occurs in derived sphenodontians (for example, *Sphenodon*, *Ankylosphenodon*, *Opisthias* and *Eilenodon*). However, there is no dental regionalization in *Priosphenodon* and there are no anterior caniniform teeth. Instead, the teeth are numerous and uniform in shape, with tooth size gradually increasing towards the posterior additional series. These teeth are rostrally pointed and closely packed, forming a continuous sharp, saw-like cutting edge (Fig. 1a, d, e). As in eilenodontines and *Opisthias*, the teeth in *Priosphenodon* are broad transversely^{11,18}, each bearing a pair of labial and lingual flanges that imbricate with the preceding tooth (Fig. 1f). Tooth and skull morphology are inter-related. *Priosphenodon* has an unusual skull, in which the dorsoventral expansion of the jugal nearly closes the lower temporal fenestra. This novel cranial architecture resists anterior torsion of the quadrate condyles¹⁹, permitting a precise orthal shearing bite and improved propaliny. The related sphenodontines achieved a similar result in a very different way, by redeveloping the lower temporal bar¹⁹. In both cases, mediolateral mobility of the jaws was restricted by extending the palatine tooth row backwards, parallel to those of the maxilla, which may have allowed a more adjusted, antero-posterior movement of the mandible (that is, eupropaliny).

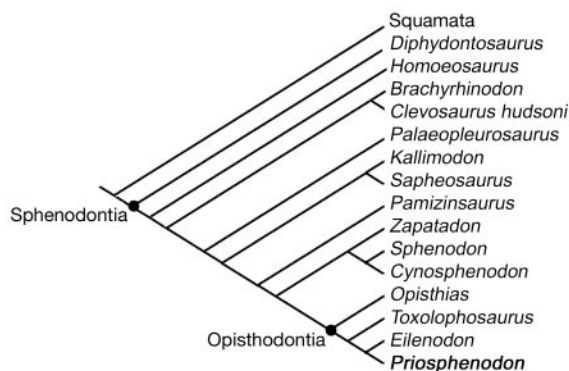


Figure 2 Cladogram depicting phylogenetic relationships of *Priosphenodon avelasi* within Sphenodontia. A data set of 67 characters and 16 taxa was analysed using Nona, resulting in a single most parsimonious tree ($L = 140$; $CI = 0.63$; $RI = 0.71$). Trees were rooted as in ref. 20. Squamata and *Diphydontosaurus avonis* were taken as the outgroup. Scoring of *Opisthias rarus* and *Eilenodon robustus* was performed with the respective holotypes and additional material from the Los Angeles County Museum, under study by S.A. (*O. rarus*: LACM 135531, 128170, 120542, 135331, 92030; *E. rasmusseni*: LACM-NCN 9). Data matrix, character list and sources, as well as algorithm information, are offered as Supplementary Information. *Priosphenodon* is included in Eilenodontinae on the basis of the following unequivocal synapomorphies: characters 41(2); 42(2); and 48(2) (short strongly upturned retroarticular process; dentary with posterior process extended until the end of the glenoid; teeth wider than they are long). They also share a very low coronoid process (reversal), and a skull length greater than 80 mm. Eilenodontinae (for example, *Toxolophosaurus*, *Eilenodon* and *Priosphenodon*) shares with *Opisthias rarus* characters 11(1); 35(3); 39(2); 45(3); 48(1); 52(1); and 58(1) (pre- and postfrontals sculptured; mandibular symphysis with a short, rounded but conspicuous anteromedial process; symmetrical glenoid facet of the lower jaw divided by medial longitudinal ridge; lack of dental regionalization; teeth subequal or mediolaterally expanded; maxillary teeth with anterolateral flanges; mandibular teeth with anteromedial flanges). They also share lingual and labial flanges that imbricate with the preceding tooth. *Priosphenodon* shares with *Eilenodon* a very large body size, with skulls larger than 100 mm.

The postcranial skeleton of *Priosphenodon* (MPCA 303) is remarkably sturdy, and the humerus is longer (with respect to the femur) than in other known sphenodontians²⁰. Moreover, the Patagonian taxon is unique among Lepidosauria in having square, distally expanded ungual phalanges (Fig. 1g, h), a condition that contrasts with the pointed unguals of most lepidosaurs^{17,19,20}.

Discussion. Character analysis (Fig. 2) places *Priosphenodon* as a close relative of the Late Jurassic *Opisthias* and the eilenodontines *Toxolophosaurus* and *Eilenodon*. These four taxa share the following: sculptured pre- and postfrontals, mandibular symphysis with a short, rounded but conspicuous anteromedial process; symmetrical glenoid facet of the lower jaw divided by medial longitudinal ridge; lack of dental regionalization; teeth subequal or mediolaterally expanded; maxillary teeth with anterolateral flanges; mandibular teeth with anteromedial flanges and lingual and labial flanges that imbricate with the preceding tooth. This set of synapomorphies supports the erection of a new taxon, Opisthodontia nov., a stem-group defined as all the sphenodonts that are more closely related to *Priosphenodon* than to *Sphenodon* (Fig. 2). The stem-based sister taxon to Opisthodontia is Sphenodontinae, defined as *Sphenodon* and all sphenodonts closer to it than to opisthodontians.

Eilenodontine opisthodontians increased in size throughout their evolutionary history. Whereas primitive terrestrial sphenodontians were small, lizard-like reptiles (a condition retained by basal sphenodontines), eilenodontines attained bulkier proportions. For example, the skulls of the Jurassic *Eilenodon* and the Early Cretaceous *Toxolophosaurus* are 60 to 100 mm in length, but in *Priosphenodon* the skull reaches 150 mm, with total adult body length slightly surpassing 1 m. *Priosphenodon* is the largest terrestrial sphenodontian yet recorded, and within opisthodontians it represents an 'extreme' of an evolutionary trend towards a large size.

Opisthodontians represent a Late Jurassic through Late Cretaceous lepidosaurian lineage that successfully inhabited South America until at least Cenomanian to Turonian times. The sister group relationships of opisthodontians and sphenodontines, coupled with the morphological features of *Priosphenodon* (for example, large adult size, stout appendicular bones, unusual ungual phalanges, sharp, beak-like structure, and bulging nasals), suggest that a hidden taxonomic diversity of Cretaceous sphenodontians remains to be documented in southern continents.

Current interpretations position post-Jurassic sphenodontians as evolutionary relicts, replaced ecologically in the Early Cretaceous by the rapidly diversifying lizards². This scenario has been drawn from the lepidosaurian fossil record of Laurasia, but it does not fit the evidence currently emerging from South America. The Mesozoic fossil record of squamates in southern continents, albeit patchy^{21–28}, provides no evidence that they were dominant over sphenodontians. The occurrence of *Priosphenodon* late in the Cretaceous, together with its numerical abundance and large body size, suggest that sphenodontians retained an important ecological role in Gondwana.

This view of South American lepidosaur faunas is further supported by the recovery of sphenodontians from the Late Campanian Los Alamos Formation of northwest Patagonia (S.A., unpublished data), showing that sphenodontians survived in this continent well into the Late Cretaceous. Although the Gondwanan record of squamates (for example, acrodontan iguanians, paramacellodid scincomorphs, putative teioids and a variety of basal forms)^{22–26} extends from the Early Jurassic onwards, modern groups of Laurasian squamates probably arrived in South America towards the end of the Campanian^{25,27,28} and dominated the small reptile fauna from the beginning of the Tertiary. In this regard, the Cenozoic record of South America resembles that of North America²⁹. Squamates did finally replace South American sphe-

dontians, but the new data from Patagonia suggests that this ecological replacement was delayed with respect to the northern continents. This may help to explain why spheodontians persisted longer in Gondwana than on other landmasses. □

Received 9 June; accepted 23 July 2003; doi:10.1038/nature01995.

1. Reynoso, V. H. A "beaded" spheodontian (Diapsida: Lepidosauria) from the Early Cretaceous of Central Mexico. *J. Vert. Paleontol.* **17**, 52–59 (1997).
2. Carroll, R. L. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
3. Novas, F. E. et al. Primer registro de lepidosaurios esfenodontes para el Cretácico Superior. *Ameghiniana* **36**, 16R (1999).
4. Apesteguía, S. et al. Los tetrapodos de 'La Buitrera', Formación Candeleros (Cenomaniano temprano), Río Negro. *I Congr. 'O. A. Reig' Vertebr. Básica Evol. Hist. Filos. Ciencia* (in the press).
5. Evans, S. E. & Sigogneau-Russell, D. New spheodontians (Diapsida: Lepidosauria: Rhynchocephalia) from the Early Cretaceous of North Africa. *J. Vert. Paleontol.* **17**, 45–51 (1997).
6. Rich, T. H. V., Molnar, R. E. & Rich, P. V. Fossil vertebrates from the Late Jurassic or Early Cretaceous Kirkwood Formation, Algoa Basin, Southern Africa. *Trans. Geol. Soc. S. Afr.* **86**, 281–291 (1983).
7. Barbera, C. & Macuglia, L. Revisione dei tetrapodi del Cretácico Inferiore di Pietraroja (Matese Orientale, Benevento) appartenenti alla collezione Costa del Museo di Paleontologia dell'Università di Napoli. *Mem. Soc. Geol. Ital.* **41**, 567–574 (1988).
8. Hugo, C. & Leanza, H. Hoja Geológica Roca (3969-IV). *Inst. Geol. Rec. Nat. SEGEMAR*, **309**, 1–95 (1999).
9. Haeckel, E. *Generelle Morphologie der Organismen. Zweiter Band: Allgemeine Entwicklungsgeschichte der Organismen* (Reimer, Berlin, 1866).
10. Williston, S. W. *The Osteology of the Reptiles* (Harvard Univ. Press, Cambridge, 1925).
11. Rasmussen, T. E. & Callison, G. A new herbivorous spheodontid (Rhynchocephalia: Reptilia) from the Jurassic of Colorado. *J. Paleontol.* **55**, 1109–1116 (1981).
12. Pol, D., Apesteguía, S., Novas, F. E. & Carignano, A. P. Evolución y paleobiogeografía de los cocodrilos del Cretácico tardío de Gondwana. *Congr. Arg. Paleontol. Bioestrat.* (in the press).
13. Gallina, P. A., Apesteguía, S. & Novas, F. E. ¿Un elefante bajo la alfombra? Los Rebbachisauridae (Sauropoda, Diplodocimorpha) del Cretácico de Gondwana. Nuevas evidencias en "La Buitrera" (Fm. Candeleros), provincia de Río Negro. *Ameghiniana* **39**, 10R (2002).
14. Apesteguía, S., Rougier, G. W., Novas, F. E. & Novacek, M. A cladotherian mammal from the Early Late Cretaceous of Patagonia. *Congr. Argent. Paleontol. Bioestrat.* (in the press).
15. Apesteguía, S. & Agnolin, F. L. Los dipnoos del Cretácico de Patagonia. *Ameghiniana* **39**, 5R (2002).
16. Fraser, N. C. & Benton, M. J. The Triassic reptiles *Brachyrhinodon* and *Polysphenodon* and the relationships of the spheodontids. *Zool. J. Linn. Soc.* **96**, 413–445 (1989).
17. Wu, X.-C. *The Comparative Anatomy and Systematics of Mesozoic Spheodontidans*. Thesis, McGill Univ. (1991).
18. Throckmorton, G. S. & Hopson, J. A. A redescription of *Toxolophosaurus cloudi* Olson, a lower Cretaceous herbivorous spheodontid reptile. *J. Paleontol.* **55**, 586–597 (1981).
19. Fraser, N. The osteology and relationships of *Clevoosaurus* (Reptilia, Spheodontida). *Phil. Trans. R. Soc. Lond. B* **321**, 125–178 (1988).
20. Reynoso, V. H. An unusual aquatic spheodontian (Reptilia: Diapsida) from the Tlayua Formation (Albian) of Central México. *J. Paleontol.* **74**, 133–148 (2000).
21. Báez, A. M. & Gasparini, Z. Orígenes y Evolución de los anfibios y reptiles del Cenozoico de América del Sur. *Acta Geol. Lilloana* **14**, 149–232 (1977).
22. Evans, S. E., Prasad, G. V. R. & Manhas, B. K. Fossil lizards from the Jurassic Kota Fm. of India. *J. Vert. Paleontol.* **22**, 299–312 (2002).
23. Zils, W., Werner, C., Moritz, A. & Saanane, C. Tendaguru, the most famous dinosaur locality of Africa. Review, survey and future prospects. *Documenta Naturae* **97**, 1–41 (1995).
24. Broschinsky, A. & Sigogneau Russell, D. Remarkable lizard remains from the Lower Cretaceous of Anoual (Morocco). *Ann. Paleontol.* **82**, 147–175 (1996).
25. Albino, A. El lagarto más antiguo de la Argentina. *I Congr. 'O. A. Reig' Vertebr. Básica Evol. Hist. Filos. Ciencia* (in the press).
26. Evans, S. E. & Yabumoto, Y. A lizard from Early Cretaceous Crato Formation, Araripe Basin, Brazil. *N. Jahrb. Paläontol. Geol.* **6**, 349–364 (1998).
27. Estes, R. & Báez, A. in *The Great American Biotic Interchange* (eds Stehli, F. & Webb, S.) 139–197 (1985).
28. Smith, K. T. Iguanian phylogeny: implications for the origin of a portion of the recent Malagasy biota. *J. Vert. Paleontol.* **21** (suppl. 3), 102A (2001).
29. Evans, S. E., Prasad, G. V. R. & Manhas, B. K. Rhynchocephalians (Diapsida: Lepidosauria) from the Jurassic Kota Formation of India. *Zool. J. Linn. Soc.* **133**, 309–334 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Avelás, Salinas and Pincheira families, C. Muñoz, L. Fernández and the ENDEMAS for their valuable help in the field. S. Reuil, A. Scaferla, M. Isasi and E. Villalba are also acknowledged for preparation of specimens; J. González and G. Lio for illustrations; and H. Canuti for photographs. Field work and study were supported by The National Geographic Society, the Agencia Nacional de Promoción Científica y Tecnológica, Akapol SA, and the Renault Argentina (to F.E.N.), and by The Jurassic Foundation and The Rotary Club (to S.A.). S. Evans, R. Molnar, G. Rougier and D. Pol substantially improved this work with useful comments and critical review.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.A. (paleoninja@yahoo.com.ar).

Ultraviolet vision in a bat

York Winter¹, Jorge López² & Otto von Helversen³

¹Department of Biology, University of Munich, Luisenstrasse 14, 80333 Munich, Germany and Max-Planck Research Centre for Ornithology, 82305 Seewiesen, Germany

²Escuela de Biología, Universidad de San Carlos de Guatemala Edificio T-10, ciudad Universitaria 01012, Guatemala

³Institute of Zoology, Erlangen University, Staudtstrasse 5, 91058 Erlangen, Germany

Most mammals, with the exception of primates, have dichromatic vision and correspondingly limited colour perception¹. Ultraviolet vision was discovered in mammals only a decade ago², and in the few rodents and marsupials where it has been found, ultraviolet light is detected by an independent photoreceptor^{2,3}. Bats orient primarily by echolocation, but they also use vision. Here we show that a phyllostomid flower bat, *Glossophaga soricina*, is colour-blind but sensitive to ultraviolet light down to a wavelength of 310 nm. Behavioural experiments revealed a spectral-sensitivity function with maxima at 510 nm (green) and above 365 nm (ultraviolet). A test for colour vision was negative. Chromatic adaptation had the same threshold-elevating effects on ultraviolet and visible test lights, indicating that the same photoreceptor is responsible for both response peaks (ultraviolet and green). Thus, excitation of the β -band of the visual pigment is the most likely cause of ultraviolet sensitivity. This is a mechanism for ultraviolet vision that has not previously been demonstrated in intact mammalian visual systems.

All bat species have functional eyes that are used in various contexts^{4–6}, but their spectral sensitivities and their capacity for spectral discrimination are not known¹. Some bat-pollinated, neotropical plant species have violet blossoms and can even reflect ultraviolet light to a remarkable degree⁷. This raised the question of whether flower bats have dichromatic colour vision and whether they perceive the ultraviolet light reflected by some flowers.

We examined these questions in three behavioural discrimination experiments with neotropical flower bats (four individuals: three female, one male) of the species *Glossophaga soricina* (Phyllostomidae). First, the detection thresholds for light stimuli at different

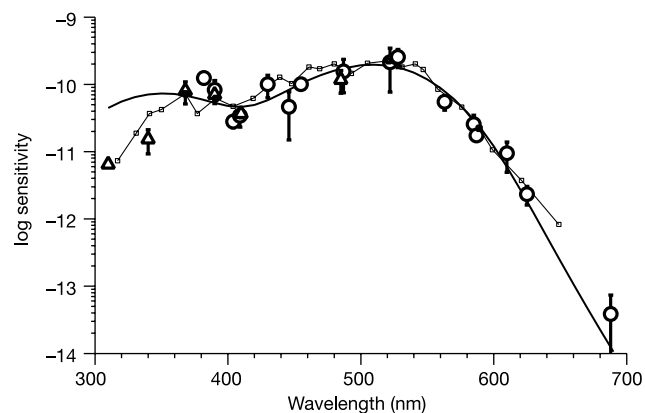


Figure 1 Spectral sensitivity function of the bat *Glossophaga soricina* (experiment I). The ordinate is the sensitivity in log units ($1/\text{number of quanta cm}^{-2} \text{ s}^{-1}$). The triangles and circles indicate the two light stimuli used (see Methods) and are means (± 1 s.d.) of the measured threshold values. The solid line is the α -band and β -band A1 pigment template with λ_{max} at 509.8 nm and β -band maximum of 349.6 nm. The line with open squares shows the scotopic spectral sensitivity of aphakic humans (who have had lenses surgically removed) measured off-fovea¹⁸ and shown here for comparison with shifted sensitivity scale.

wavelengths, including the ultraviolet range, were determined. The result was a bimodal spectral-sensitivity function (Fig. 1). This led us to investigate the ability of *G. soricina* to discriminate between colours, which as a property of the whole system can be tested only by psychophysical experiments. In a final experiment, the underlying photoreceptor mechanism was examined by using spectral chromatic adaptation. The result proves that these bats are capable of ultraviolet perception through a single receptor mechanism not previously demonstrated in mammals with intact eyes.

For all experiments, *G. soricina* bats were operantly conditioned under scotopic conditions in a two-alternative forced-choice paradigm to discriminate between two test panels, one (experiments I and III) or two (experiment II) of which presented a light stimulus with adjustable wavelength and intensity. In experiment I, the spectral sensitivity (the reciprocal of the threshold intensity) of *G. soricina* extended from 310 nm to 688 nm (Fig. 1). The two sensitivity maxima found are in the ultraviolet region at a wavelength above 365 nm, and in the green, at 510 nm (Fig. 1). Between individuals and between the two sexes, the spectral sensitivities did not differ in their degree of fit between measurement and the photopigment template estimated from those data (three-way analysis of variance (ANOVA), $F_{1,43} = 1.10$, $P = 0.30$). Further experiments were performed with female individuals. The possibility that ultraviolet sensitivity was produced by a second, separate photoreceptor was tested by measuring the spectral sensitivity under chromatic adaptation (see experiment III).

Can bats distinguish spectral colours? In experiment II, a conditioned light stimulus (at 382 nm) had to be distinguished from two test wavelengths: green (522 nm) or orange (585 nm). Stimuli were presented at constant intensities of 2.5 log units above their respective sensitivity thresholds so that stimulus pairs could not be discriminated on the basis of brightness. As a sensitive test of learning, we evaluated the expected increase in correct choices with number of test experiences. Even after several thousand choices, no effect of learning was discernable. The number of correct choices did not increase with the number of test experiences (382 versus 522: Pearson's correlation, $r = -0.002$, $P = 0.88$, $n = 8,247$; 382 versus 585: Pearson's correlation, $r = 0.009$, $P = 0.53$, $n = 5,176$). It follows that *G. soricina* is colour-blind (Fig. 2).

The bimodal threshold curve raised the question of whether the response peak in the ultraviolet range is attributable to a separate visual pigment. This was addressed in experiment III by threshold

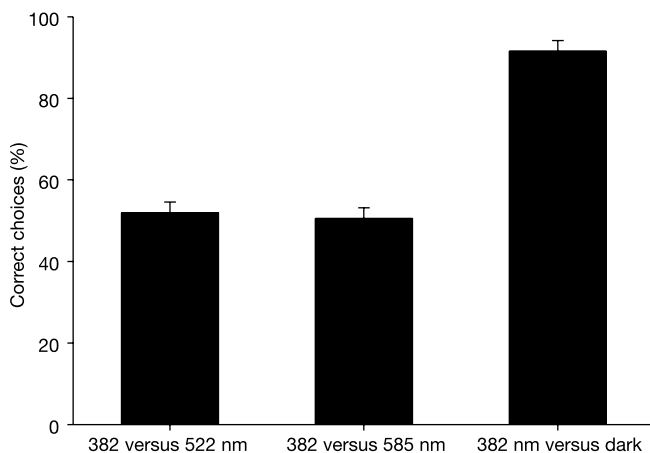


Figure 2 Test for colour vision (experiment II). In a two-alternative forced-choice paradigm a bat (F2) was first conditioned to a 382-nm light stimulus. After task acquisition, the bat was tested to distinguish between stimuli of 382 and 522 nm and of 382 and 585 nm. Data are means ($\pm$ 1 s.d.) from different experimental sessions (522 nm: 15/8,210; 585 nm: 8/5,179; 382 nm versus dark: 15/7,610; number of sessions per total number of decisions).

measurements under selective chromatic adaptation. The adaptation of a receptor leads to a decrease in its sensitivity function. Receptors with different response peaks should respond to the same adaptation wavelength differently. Here, we measured detection thresholds at two wavelengths (382 nm and 522 nm), but now with a simultaneous, continuous background adaptation light. The results showed that selective chromatic adaptation in the wavelength regions containing the two maxima did not produce any significant, independent change in the sensitivity in these two regions when tested with different adaptation wavelengths (Fig. 3). Hence, the ratio of the sensitivities in the two peak regions remained constant even when different chromatic adaptation lights were used. This result proves that the spectral sensitivities in the regions of the two maxima are not independent of one another under chromatic adaptation. Instead, they derive from the activity of a single receptor.

The ability of the flower-visiting bat species *G. soricina* to detect ultraviolet light up to a wavelength of at least 310 nm is the first demonstration of ultraviolet vision in a mammal other than rodents and marsupials. Ultraviolet light sensitivity is probably a characteristic of a relatively small species. Imaging errors caused by chromatic aberration in the ultraviolet region⁸ increase with the size of the eye, as does Rayleigh scattering⁹, which would be particularly disturbing for larger species. Furthermore, for colour systems sensitive to ultraviolet light, it might be relatively difficult to maintain colour constancy in a fluctuating illumination spectrum^{10,11}. Consistent with this, *G. soricina* has ultraviolet vision but is colour-blind, and previous findings of ultraviolet sensitivity have been restricted to small, night-active mammals^{2,12}. The possible adaptive function of ultraviolet vision in this species may be to enhance visual contrast perception¹³ of ultraviolet-light-reflecting flowers⁷ during the twilight phase of the day, when the spectrum is shifted towards short wavelengths¹⁴. This would increase flower-search efficiency.

Ultraviolet vision in rodents is based on a dichromatic visual system with an independent receptor for ultraviolet light detection^{3,15}. However, the threshold changes observed under chromatic

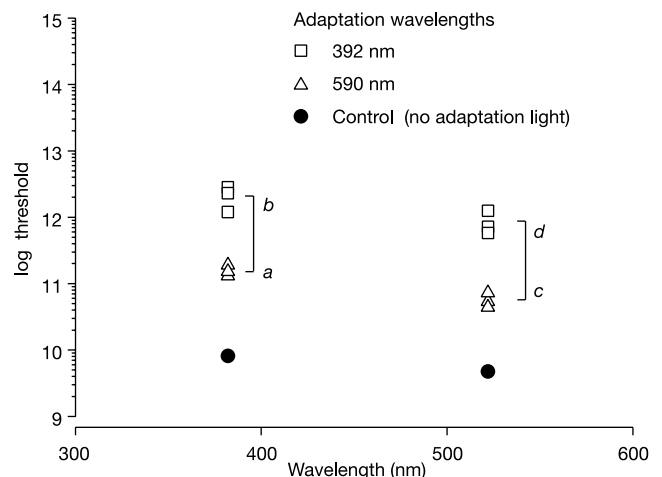


Figure 3 Displacement of sensitivity thresholds under chromatic adaptation (experiment III). Sensitivity thresholds at 382 nm and at 522 nm were measured under two different conditions of chromatic adaptation and also in darkness (control). The thresholds *a*, *b* and *c*, *d* were shifted with respect to the control threshold determined in darkness as a result of chromatic adaptation. The ratios $a/b (\pm$ s.d.) ($1.57 \times 10^{11} \pm 0.14 \times 10^{11} / 2.11 \times 10^{12} \pm 0.83 \times 10^{12} = 0.074$) and $d/c (\pm$ s.d.) ($5.67 \times 10^{10} \pm 1.43 \times 10^{10} / 8.50 \times 10^{11} \pm 0.14 \times 10^{11} = 0.067$) did not significantly differ from one another (ANOVA, $F_{1,4} = 0.02$, $P = 0.89$). The ordinate is log light intensity (number of quanta $\text{cm}^{-2} \text{s}^{-1}$) at which the threshold criterion was reached.

adaptation in bats in the present study suggest the opposite conclusion: that the visual system of *Glossophaga* is monochromatic and does not include a separate ultraviolet light receptor. Because the absorption range of opsin-based pigments extends into the ultraviolet region¹⁶, ultraviolet vision does not necessarily require the presence of a specific receptor. However, because ocular media act as ultraviolet light filters¹⁷, use of this β -band sensitivity was not expected in mammals. Nevertheless, ultraviolet light perception of *Glossophaga* in all likelihood is accomplished by stimulation of the β -band of the photoreceptor. This is a mechanism of ultraviolet vision that has not previously been shown in intact eyes of mammals. Exceptions are aphakic humans (people who have had lenses surgically removed). Their rod-based spectral sensitivity closely resembles the spectral sensitivity obtained here¹⁸ (Fig. 1, open squares). Thus, ultraviolet light absorbance by intact ocular media of the small eye of *G. soricina* is equal to ultraviolet absorbance by the cornea and by aqueous and vitreous humour in the absence of lenses in the larger human eye. The close correspondence of scotopic spectral sensitivities between bat and aphakic humans corroborates our inference of a single photoreceptor mechanism underlying ultraviolet and scotopic visual sensitivity in *Glossophaga* bats. □

Methods

Bats orient by echolocation, and therefore experiments were conducted in darkness without ambient adaptation light. Animals were adapted to scotopic conditions for at least 1 h before measurements were taken. Choice of the non-illuminated stimulus was always rewarded in experiments I and III from an artificial flower underneath the respective test panel.

Experiment I

Spectral sensitivities were determined on the basis of 67 threshold measurements encompassing more than 75,000 behavioural decisions by three bats (females F1 and F3, and male M4). The pigment template in Fig. 1 was calculated from equations 1 and 4, and Fig. 7 from ref. 19, and the data from the two females. The best fit was determined by a least-squares method. The degree of fit between measurements and template was compared by three-way ANOVA with individual, sex and wavelength as independent variables, and log difference between measurement and template as the dependent variable. In the spectral range from 310 nm to 485 nm, light stimuli were produced with a xenon high-pressure lamp (150 W, XBO150, Osram; VX 150/1 kf-2, Siemens AG) and guided through quartz light conductors (active diameter 5 mm) to the test panels (triangles in Fig. 1). The intensity was adjusted by quartz glass filters. For the spectral range from 382 nm to 688 nm, we used 5-mm light-emitting diodes (Kingbright Electronics; Marl International Ltd) inserted directly into the test panels (circles in Fig. 1). The spectral composition of the light stimuli was produced or modified using interference filters (Schott). The stimulus intensity was determined by using a calibrated photomultiplier with a gallium arsenide photoelectric cell (type PM270C, detector PM-270-OD, International Light). The measured values of stimulus intensity in W cm^{-2} were corrected with the spectral-sensitivity curve of the photomultiplier and converted to number of quanta $\text{cm}^{-2} \text{s}^{-1}$. This value is given by the energy of a stimulus divided by the energy of a light quantum (Q) (with $Q = hc/\lambda$; $h = 6.626176 \times 10^{-34} \text{ J s}^{-1}$ (Planck quantum of action); $n = c/\lambda$; $c = 3 \times 10^8 \text{ m s}^{-1}$ (speed of light); λ = wavelength). The spectral emission curves of all stimuli and adaptation lights (see experiment III) were determined using a photometer (Spex 1700-III) with photomultiplier (S-20 R562, Hamamatsu).

Experiment III

The adaptation light used was a $10 \times 10 \text{ cm}$ illuminated panel (light box). The stimulus light (382 nm or 522 nm) was positioned within a 5-mm hole in the centre of this light box and was thus surrounded by adaptation light. Adaptation lights were generated in the ultraviolet light range ($\lambda_{\text{max}} = 392 \text{ nm}$) by UG-1 glass filters (Schott) as panels in front of four fluorescent lamps (Linux-9W/78, Radium Wipperfurth), and within the red range ($\lambda_{\text{max}} = 590 \text{ nm}$) by RG glass filter panels (Schott) in front of incandescent lamps. Data based on 16 threshold measurements (three each for a , b , c and d , and four controls at 382 nm and 522 nm; see Fig. 3) were obtained from individual F1.

Received 14 May; accepted 24 July 2003; doi:10.1038/nature01971.

- Jacobs, G. H. The distribution and nature of colour vision among the mammals. *Biol. Rev.* **68**, 413–471 (1993).
- Jacobs, G. H., Neitz, J. & Deegan, J. F. II Retinal receptors in rodents maximally sensitive to ultraviolet light. *Nature* **353**, 655–656 (1991).
- Jacobs, G. H. & Deegan, J. F. II Sensitivity to ultraviolet light in the gerbil (*Meriones unguiculatus*): Characteristics and mechanisms. *Vision Res.* **34**, 1433–1441 (1994).
- Suthers, R., Chase, J. & Bradford, B. Visual form discrimination by echolocating bats. *Biol. Bull.* **137**, 535–546 (1969).
- Chase, J. Differential responses to visual and acoustic cues during escape in the bat *Anoura geoffroyi*: cue preferences and behavior. *Anim. Behav.* **31**, 526–531 (1983).

- Winter, Y. Flight speed and body mass of nectar-feeding bats (Glossophaginae) during foraging. *J. Exp. Biol.* **202**, 1917–1930 (1999).
- Biedinger, N. & Barthlott, W. Untersuchungen zur Ultraviolettreflexion von Angiospermenblüten. I. Monocotyledoneae. *Trop. Subtrop. Pflanzenwelt* **86**, 1–122 (1993).
- Neumeyer, C. *Das Farbsehen des Goldfisches* (Georg Thieme, Stuttgart, New York, 1988).
- Kirschfeld, K. Carotenoid pigments: their possible role in protecting against photooxidation in eyes and photoreceptor cells. *Proc. R. Soc. Lond. B* **216**, 71–82 (1982).
- von Campenhausen, C. Photoreceptors, lightness constancy and color vision. *Naturwissenschaften* **73**, 674–675 (1986).
- Dyer, A. G. Broad spectral sensitivities in the honeybee's photoreceptors limit color constancy. *J. Comp. Physiol. A* **185**, 445–453 (1999).
- Arrese, C. A., Hart, N. S., Thomas, N. M., Beazley, L. D. & Shand, J. Trichromacy in Australian marsupials. *Curr. Biol.* **12**, 657–680 (2002).
- Remy, M. & Emmerton, J. Behavioral spectral sensitivities of different retinal areas in pigeons. *Behav. Neurosci.* **103**, 107–177 (1989).
- Smith, H. Light quality, photoperception and plant strategy. *Annu. Rev. Plant Physiol.* **33**, 481–518 (1982).
- Jacobs, G. H. Ultraviolet vision in vertebrates. *Am. Zool.* **32**, 544–554 (1992).
- Morton, R. A. in *Handbook of Sensory Physiology VIII/I* (ed. Dartnall, H. J. A.) 33–68 (Springer, Berlin, 1972).
- Douglas, R. H. & Marshall, N. J. in *Adaptive Mechanisms in the Ecology of Vision* (eds Archer, S. N., Djamgoz, M. B. A., Loew, E. R., Partridge, J. C. & Valleria, S.) 95–162 (Kluwer, New York, 1999).
- Griswold, S. M. & Stark, W. S. Scotopic spectral sensitivity of phakic and aphakic observers extending into the near ultraviolet. *Vision Res.* **32**, 1739–1743 (1992).
- Govardovskii, V. I., Fyhrquist, N., Reuter, T., Kuzmin, D. G. & Donner, K. In search of the visual pigment template. *Vis. Neurosci.* **17**, 509–528 (2000).

Acknowledgements We thank C. Peppermüller and H. Sadowski for stimulus calibrations, and G. Seitz for general support. This work was supported by a grant from the Volkswagen Foundation to Y.W. J.E.L. was supported by the German Academic Exchange Service (DAAD).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.W. (winter@zi.biologie.uni-muenchen.de).

Consolidation during sleep of perceptual learning of spoken language

Kimberly M. Fenn¹, Howard C. Nusbaum¹ & Daniel Margoliash^{1,2}

¹Department of Psychology and ²Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, Illinois 60637, USA

Memory consolidation resulting from sleep has been seen broadly: in verbal list learning¹, spatial learning^{2,3}, and skill acquisition in visual^{4–8} and motor^{9–11} tasks. These tasks do not generalize across spatial locations or motor sequences, or to different stimuli in the same location^{5,11,12}. Although episodic rote learning constitutes a large part of any organism's learning, generalization is a hallmark of adaptive behaviour¹³. In speech, the same phoneme often has different acoustic patterns depending on context. Training on a small set of words improves performance on novel words using the same phonemes but with different acoustic patterns, demonstrating perceptual generalization¹⁴. Here we show a role of sleep in the consolidation of a naturalistic spoken-language learning task that produces generalization of phonological categories across different acoustic patterns. Recognition performance immediately after training showed a significant improvement that subsequently degraded over the span of a day's retention interval, but completely recovered following sleep. Thus, sleep facilitates the recovery and subsequent retention of material learned opportunistically at any time throughout the day. Performance recovery indicates that representations and mappings associated with generalization are refined and stabilized during sleep.

In perceptual learning of synthetic speech, listeners are presented with speech produced by a computer-controlled text-to-speech synthesizer. Synthetic speech is difficult to understand and naive listeners make a greater number of perceptual errors than they do in recognizing natural speech, but show significant learning, improving by an average of 45 percentage points after eight 1-h daily training sessions. Once acquired, performance improvement lasts for at least six months¹⁴. This magnitude of learning is all the more remarkable in that listeners never hear the same word twice—all learning is generalization^{14,15}. This paradigm allowed us to investigate the role of sleep in a naturalistic learning task that produces generalization. Two groups were given a pretest, trained, and then retested after a 12-h retention period. For one group, pretest and training started at 9 a.m., and the post-test was given after 12 h of waking, at 9 p.m.; for the second group, pretest and training were at 9 p.m., and the post-test was given at 9 a.m. the following morning, after 12 h that included sleep. A third, control group received pretest, training and post-test within a single session with no retention interval.

Training had a strong, reliable and immediate effect on the perception of synthetic speech. Control-group identification accuracy ($n = 24$ participants) increased by an average of 21.4 ± 1.6 (mean $\pm$ s.e.m.) percentage points from pretest (32.9 ± 1.9) to post-test (54.3 ± 2.5). The improved performance represents generalization, because the pretest, post-test and training words were all different. By comparison, after a 12-h waking period, performance accuracy improved by only 10.1 ± 2.0 percentage points between pretest (37.2 ± 2.4) and post-test (47.3 ± 3.3) ($n = 12$), a significant reduction compared with the full control group (Fisher's $P < 0.01$) and by comparison with the matched morning control subgroup ($t_{22} = 2.6$, $P < 0.02$), showing that the two groups, trained at the same time, differ significantly after a 12-h waking retention period. After a regular sleep period, however, performance improved by an average of 18.7 ± 1.6 points ($n = 12$). Learning displayed by the 12-h sleep group did not differ from that of the control group ($P = 0.28$), but both groups performed significantly better than the 12-h waking retention group (Fisher's $P < 0.01$ for both) (Fig. 1). We also confirmed these results within a single group ($n = 12$), given a pretest and training in the morning, a post-test after a 12-h waking period, and a second post-test after a

12-h sleep period. Performance improved by 10.5 ± 2.6 percentage points from morning pretest (28.3 ± 1.9) to evening post-test (38.9 ± 3.2), and significantly more (19.1 ± 2.0 percentage points, $P < 0.01$) from pretest to the following morning's post-test (47.6 ± 2.4).

The performance difference after 12 h awake and 12 h including regular sleep might be due to diurnal effects on testing or training. However, a comparison of morning pretest performance (32.8 ± 1.3) with evening pretest performance (32.7 ± 1.7) for all groups revealed no reliable diurnal difference ($F_{1,82} = 0.005$, $P = 0.94$). Furthermore, half of the control group was tested and trained in the morning, and the other half in the evening. The evening group ($n = 12$) improved by an average of 24.3 ± 1.5 points, whereas the morning group improved by an average of 18.5 ± 2.6 points, a marginally significant effect ($P = 0.06$). Indeed, young adults are at the acrophase of their circadian cycle in the evening, and perform better at short-term memory and other cognitive tasks than in the morning¹⁶, a circadian effect both opposite in direction and much smaller in magnitude than the performance deficit we observed. Thus, simple circadian effects on pre- or post-test performance cannot explain the current findings.

Learning (in contrast to test performance) may also exhibit circadian effects. To examine this possibility, two further groups were tested, receiving pretest and training either in the morning ($n = 12$) or in the evening ($n = 12$). The post-test was given after a 24-h retention period, controlling for circadian effects on test performance. If sleep stabilizes learning associated only with the performance level achieved immediately before sleep, performance for the 24-h retention groups should be predicted by performance levels at the time of sleep (high for the evening-trained and low for the morning-trained groups). Instead, participants showed learning of 18.4 ± 2.1 and 18.9 ± 1.6 percentage points for morning and evening testing, respectively. These groups did not differ significantly from each other ($P = 0.86$) or from the 12-h sleep group ($P = 0.93$ for both), but both differed significantly from the 12-h awake group ($P < 0.01$ for both comparisons). Clearly, performance levels were independent of when the training took place during the first day. Moreover, sleep stabilized learning so that subsequent waking did not adversely affect retention of learning by the following evening.

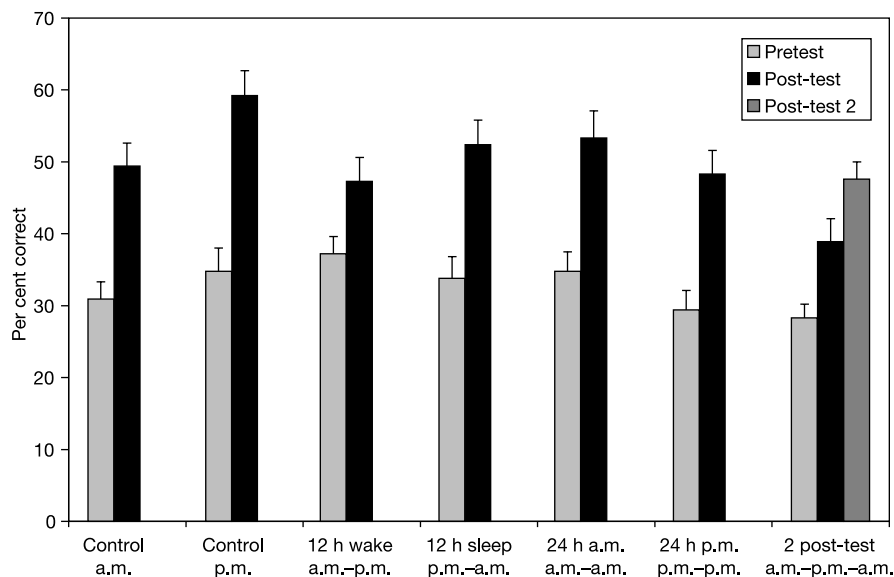


Figure 1 Sleep effect on retention of learning for word identification. Pairs of bars indicate pretest (light grey) and post-test (black) scores for each condition ($\pm$ s.e.m.). Starting with the leftmost pair of bars: 9 a.m. control group, 9 p.m. control group, 12-h waking group,

12-h sleep-phase group, 24-h group with 9 a.m. sessions, 24-h group with 9 p.m. sessions, 24-h group with 9 p.m. and 9 a.m. post-tests (dark grey bar is post-test 2 at 9 a.m.).

We have demonstrated consolidation during sleep for complex, generalized skill acquisition. Previous research has shown that sleep affects perceptual^{4,7,8} and motor learning¹⁰, in tasks limited to specific patterns or spatial locations. In the present study, participants learned a new mapping from complex acoustical patterns to pre-existing linguistic categories, which generalized to new stimuli^{14,15}. This behaviour involves distinct processes¹⁷, the formation of specific memories associated with the learned words (episodic, declarative representations), and the establishment of a mapping defined over the set of learned words that supports generalization to new utterances (procedural learning). When participants are given equal amounts of training across days with a small set of repeated words or with entirely novel words, different patterns of learning are seen. Training on a set of repeated words produces near-perfect performance on those words, but very poor generalization to novel test items. Training on all novel items produces a much larger generalization effect on the post-test¹⁵. The generalization effect cannot be accounted for by memorizing acoustic patterns of phonemes, because different acoustic patterns may represent the same phoneme and the same pattern may represent different phonemes¹⁸, depending on context. These context-conditioned effects even span syllable boundaries¹⁹ and are used in perception²⁰. The acoustic patterns of phonemes in different phonetic contexts cannot be statistically inferred from the distribution of a sample of those utterances²¹, and rote memorization and linear interpolation across a small set of acoustic patterns for each consonant and vowel cannot explain human speech recognition²². To recognize new words, listeners must learn to generalize, predict the acoustic consequences of different phonetic contexts.

Sleep has at least two separate effects on learning. Sleep consolidates memories, protecting them against subsequent interference or decay. Sleep also appears to 'recover' or restore memories. In the perceptual learning task we used, memories are sufficiently robust to last for up to six months¹⁴ or, in a comparable reading-acquisition task, even a year²³. Such robust memory represents a significant selective advantage that an organism might accrue from sleep-mediated processes. Learning can take place at any time during a waking period and any loss due to decay or interference will be restored by sleep. This also implies that a selective advantage of sleep is to enable organisms to learn opportunistically any time during the day without penalty as to robustness of learning.

We do not know if the reduction in performance observed after periods of wakefulness is due to decay of learned material, or to interference from listening to speech or other cognitive processing during the day. If performance is reduced by interference, sleep might strengthen relevant associations and/or weaken irrelevant associations, improving access to relevant memories. If performance is reduced by decay, sleep might actively recover what has been lost, presumably by an interaction between partially retained memories (words) and partially retained mappings that resulted from learning the word set. □

Methods

Participants listened to computer-generated monosyllabic consonant-vowel-consonant (CVC) words taken from a phonetically balanced (PB) list (approximating the distribution of phonemes in English)²⁴, and responded by typing the word. During training, a series of synthetic speech words were presented over headphones paired with the printed form of the word as feedback. After each training block, participants identified the trained words. A pretest and post-test were given before and after training, during which participants identified different sets of words without feedback. The pretest and post-test each consisted of 100 PB words, and the two training sessions each consisted of 150 PB words. The training sessions were structured into three blocks of 50 words. Participants rested between blocks. Word lists for testing and training were counterbalanced across participants. The 84 participants (all groups) were 20.3 ± 2.3 (mean $\pm$ s.d.) years old; each participated in one group only. Stimuli were delivered through Sennheiser HD 570 headphones with an r.m.s. sound pressure level of 66.5 dB.

One group had two post-training periods of testing. In the second post-test, participants were tested with a further set of 100 PB words that no other group was trained or tested on. The consistency of the result for the second post-test argues against any materials effect and further emphasizes the robustness of the findings.

Received 3 June; accepted 24 July 2003; doi:10.1038/nature01951.

1. Ficca, G., Lombardo, P., Rossi, L. & Salzarulo, P. Morning recall of verbal material depends on prior sleep organization. *Behav. Brain Res.* **112**, 159–163 (2000).
2. Smith, C. & Rose, G. Posttraining paradoxical sleep in rats is increased after spatial learning in the Morris water maze. *Behav. Neurosci.* **111**, 1197–1204 (1997).
3. Plihal, W. & Born, J. Effects of early and late nocturnal sleep on priming and spatial memory. *Psychophysiol.* **36**, 571–582 (1999).
4. Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M. & Sagi, D. Dependence on REM sleep of overnight perceptual skill. *Science* **265**, 679–682 (1994).
5. Karni, A. & Sagi, D. The time course of learning a visual skill. *Nature* **365**, 250–252 (1993).
6. Mednick, S. C. *et al.* The restorative effect of naps on perceptual deterioration. *Nature Neurosci.* **5**, 677–681 (2002).
7. Stickgold, R., James, L. & Hobson, J. A. Visual discrimination learning requires sleep after training. *Nature Neurosci.* **3**, 1237–1238 (2000).
8. Stickgold, R., Whidbee, D., Schirmer, B., Patel, V. & Hobson, J. A. Visual discrimination task improvement: A multi-step process occurring during sleep. *J. Cogn. Neurosci.* **12**, 246–254 (2000).
9. Brashers-Krug, T., Shadmehr, R. & Bizzi, E. Consolidation in human motor memory. *Nature* **382**, 252–255 (1996).
10. Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A. & Stickgold, R. Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron* **35**, 205–211 (2002).
11. Karni, A. *et al.* The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proc. Natl Acad. Sci. USA* **95**, 861–868 (1998).
12. Karni, A. *et al.* Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature* **377**, 155–158 (1995).
13. Karni, A. & Bertini, G. Learning perceptual skills: behavioral probes into adult cortical plasticity. *Curr. Opin. Neurobiol.* **7**, 530–535 (1997).
14. Schwab, E. C., Nusbaum, H. C. & Pisoni, D. B. Some effects of training on the perception of synthetic speech. *Hum. Factors* **27**, 395–408 (1985).
15. Greenspan, S. L., Nusbaum, H. C. & Pisoni, D. B. Perceptual learning of synthetic speech produced by rule. *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 421–433 (1988).
16. Johnson, M. P. *et al.* Short-term memory, alertness and performance: a reappraisal of their relationship to body temperature. *J. Sleep Res.* **1**, 24–29 (1992).
17. Gabrieli, J. D. E. Cognitive neuroscience of human memory. *Annu. Rev. Psychol.* **49**, 87–115 (1998).
18. Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
19. Ohman, S. E. Coarticulation in VCV utterances: Spectrographic measurements. *J. Acoust. Soc. Am.* **39**, 151–168 (1966).
20. Rudnicki, A. I. & Cole, R. A. Effect of subsequent context on syllable perception. *J. Exp. Psychol. Hum. Percept. Perform.* **4**, 638–647 (1978).
21. Nusbaum, H. C. & Pisoni, D. B. Automatic measurement of speech recognition performance: A comparison of six speaker-dependent recognition devices. *Comp. Speech Lang.* **2**, 87–108 (1987).
22. Pisoni, D. B. in *Handbook of Learning and Cognitive Processes Vol. 6 Linguistic Functions in Cognitive Theory* (ed. Estes, W. K.) 167–233 (Lawrence Erlbaum Assoc., Hillsdale, New Jersey, 1978).
23. Kollers, P. A. & Perkins, D. N. Orientation of letters and errors in their recognition. *Percept. Psychophys.* **5**, 265–269 (1969).
24. Egan, J. P. Articulation testing methods. *Laryngoscope* **58**, 955–991 (1948).

Acknowledgements We thank E. Van Cauter, A. S. Henly and N. P. Issa for critiques of the manuscript. This work was supported by grants from the McCormick Tribune Foundation to H.C.N. and from the National Institutes of Health to D.M.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.M.F. (kmfenn@uchicago.edu).

Dissociable stages of human memory consolidation and reconsolidation

Matthew P. Walker, Tiffany Brakefield, J. Allan Hobson & Robert Stickgold

Laboratory of Neurophysiology, Massachusetts Mental Health Center, Department of Psychiatry, Harvard Medical School, 74 Fenwood Road, Boston, Massachusetts 02115, USA

Historically, the term 'memory consolidation' refers to a process whereby a memory becomes increasingly resistant to interference from competing or disrupting factors with the continued passage of time¹. Recent findings regarding the learning of skilled sensory and motor tasks ('procedural learning') have refined this definition, suggesting that consolidation can be more strictly

determined by time spent in specific brain states such as wake, sleep or certain stages of sleep²⁻⁸. There is also renewed interest⁹ in the possibility that recalling or 'reactivating' a previously consolidated memory renders it once again fragile and susceptible to interference¹⁰⁻¹², therefore requiring periods of reconsolidation¹³⁻¹⁵. Using a motor skill finger-tapping task, here we provide evidence for at least three different stages of human motor memory processing after initial acquisition. We describe the unique contributions of wake and sleep in the development of different forms of consolidation, and show that waking reactivation can turn a previously consolidated memory back into a labile state requiring subsequent reconsolidation.

We have previously shown that following practice of a specific motor sequence, delayed performance improvements only occur across a night of sleep, while waking periods of 4, 8 or even 12 hours offer no such performance enhancements^{7,8}. In the present work, we studied eight groups of subjects using the same motor skill finger-tapping task (Fig. 1). Subjects in group 1 demonstrated sleep-dependent enhancements in performance speed (Fig. 2a, right) and accuracy (Fig. 2a, left) when tested 24 hours after training, similar to earlier findings^{7,8}. However, when subjects in group 2 were trained on a second motor sequence immediately after the first, interference was seen, so that overnight improvement in accuracy only occurred for the second sequence, and not the first (Fig. 2b). Significant overnight improvement in speed was observed for both sequences.

It was possible that the second interference sequence immediately reversed the initial learning achieved during training on the first sequence. This was not the case, however, as subjects in group 3 showed no loss of speed or accuracy for either the first or second motor sequence when retested immediately after learning the second interference sequence, rather than after 24 hours (Fig. 2c). Thus, the interference effect seen in group 2 appeared to reflect a specific disruption of the subsequent consolidation process.

Although it was clear that 4–12 hours of waking did not enhance behavioural performance⁷, we did not know whether this waking time period could stabilize the motor memory. Subjects in group 4 were therefore trained on the first motor sequence at 10 a.m., but learned the second, interference sequence 6 hours later, at 4 p.m. In contrast to subjects in group 2, who underwent immediate interference training, subjects in group 4 now showed significant additional overnight gains in both speed and accuracy (Fig. 3). Thus it appears that 6 hours of waking can stabilize, but not enhance, this type of motor memory; the enhancement phase of consolidation requires sleep.

To determine the effects of post-training periods in excess of 24 hours, subjects in group 5 were trained only on one motor sequence on day 1 and retested 24 and 48 hours later (Fig. 4a). When retested after 24 hours, normal improvement in speed and accuracy was seen. At 48 hours, both speed and accuracy improved again, although the increase in accuracy was not significant. This is consonant with our previous findings⁸ that showed similar improvement in a group tested after 72 hours. Thus, there is improvement, rather than decay, of motor memory after a second and third post-training night of sleep.

As expected, when subjects in group 6 were trained on the first sequence on day 1, and learned a second sequence, 24 hours later, on day 2, performance for the first sequence suffered no effect of interference, improving significantly by day 3 (Fig. 4b). Therefore, the stabilization seen 6 hours after training is subsequently maintained for at least 24 hours.

But, to our surprise, this stabilization could be reversed. In contrast to group 6, subjects in group 7 (Fig. 4c) briefly rehearsed the first sequence immediately before learning the second, interference sequence on day 2, demonstrating improvement in both speed and accuracy. However, when retested again on the first sequence on day 3, learning had been reversed, with performance

accuracy decreasing significantly, by more than 50%, and speed showing a small but non-significant reduction. At the same time, performance on the second, interference sequence showed significant overnight improvements in both speed and accuracy, similar to normal 24 hours improvement. The performance decreases for the first sequence on day 3 were significantly different compared to both the initial improvements achieved on this sequence on day 2 (+24 hours) and the improvements for the second sequence on day 3 (+24 hours). Furthermore, the loss of learning observed in group 7 on day 3 was also significantly different compared to the increases in performance for the first sequence on day 3 in groups 5 and 6 (accuracy, $P < 0.030$; speed, $P < 0.006$) (Fig. 4a–c). Thus, recollec-

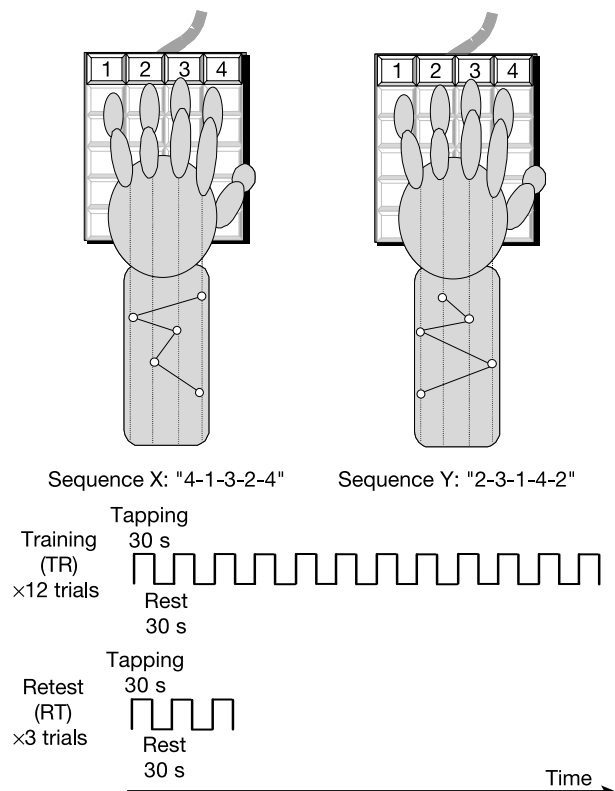


Figure 1 Procedural motor skill task and experimental protocol. One hundred right-handed subjects, aged 18–27 (s.d. ± 2.1) and without neurological, psychiatric or sleep-disorder histories, were assigned to eight experimental groups (see Figs 2–4). Subjects performed a sequential finger tapping motor skill task (for detailed methods see ref. 7). The task involves pressing four numeric keys using the fingers of the left (non-dominant) hand, repeating a five-element sequence "as quickly and accurately as possible" for a trial period of 30 s, followed by 30 s of rest. Two motor patterns containing completely unique grammars (sequence X: 4-1-3-2-4; sequence Y: 2-3-1-4-2) were used in a balanced order as either the first or second sequence learned. Subjects were also retested on the first or second sequence in a balanced order to eliminate potential retest-ordering effects. Training (TR) consisted of 12 contiguous trials, while retesting (RT) involved three contiguous trials. Performance measures were the number of complete sequences achieved ('speed'), and the number of errors made relative to the number of correct sequences ('accuracy'). Comparisons within each experimental group were performed using repeated measures analysis of variance (ANOVA) and paired *t*-tests (two-tailed). Comparisons between groups were performed using unpaired *t*-tests (two-tailed). Retest values were compared to the final three trials of training, allowing true evaluation of between session, time-dependent learning, instead of the mean training value across all 12 trials. There were no significant differences between the initial learning curves during training for the first sequence compared to the second sequence learned in all groups that performed two sequences (average of groups 2–6 and 8; speed: ANOVA, $P = 0.423$; accuracy: ANOVA, $P = 0.408$).

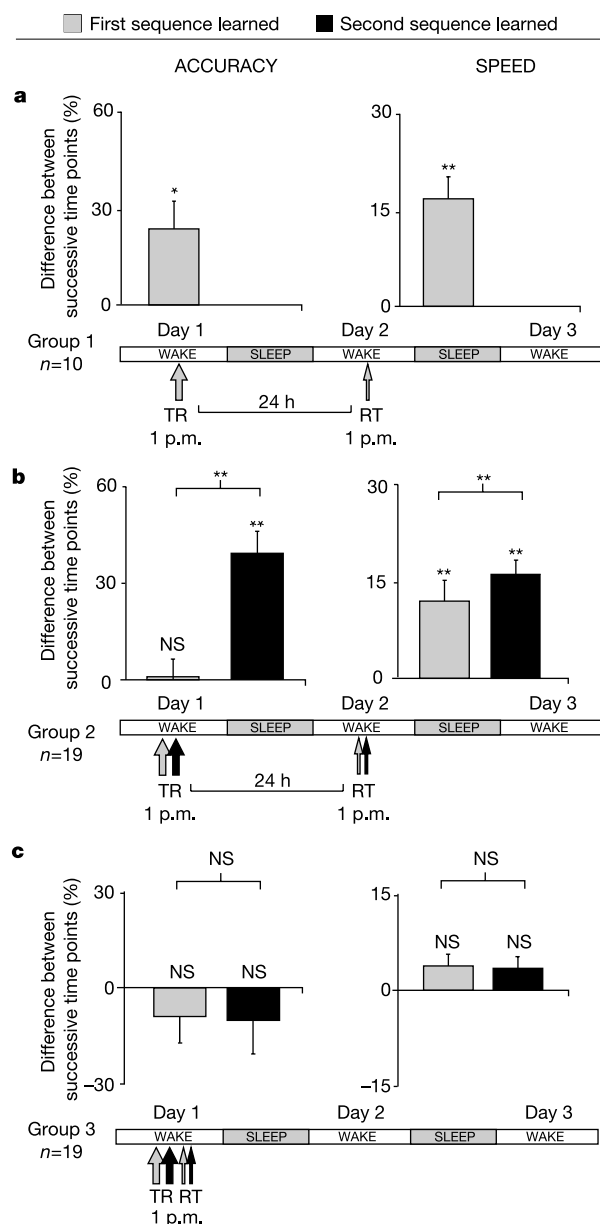


Figure 2 Changes in motor learning in groups 1–3. **a**, Group 1: following training on a single motor sequence on day 1 (thick arrow; TR), overnight increases in performance speed ($P < 0.001$) and accuracy ($P = 0.023$) were seen at the 24-hour retest on day 2 (thin arrow; RT). **b**, Group 2: immediately after learning the first motor sequence on day 1, subjects now learned a second motor sequence. When retested on day 2, improvements in accuracy occurred only for the second sequence ($P < 0.001$; filled black bar), while no such improvement developed for the first sequence ($P = 0.90$; grey bar). Improvements in speed ($P < 0.001$) were observed for both sequences. **c**, Group 3: subjects were again trained on the second motor sequence directly after the first. When retested after 5 min, performance indicated retention of initial training improvements for both sequences, but without the improvements in speed or accuracy seen 24 hours later. Error bars are s.e.m. Asterisks represent significance (P): * ≤ 0.05 ; ** ≤ 0.005 ; NS, non-significant. Same for all subsequent figures.

tion or reactivation of the first sequence memory during retesting on day 2 returned it to a labile state, making it once more susceptible to interference.

To confirm that interference following reactivation specifically disrupted subsequent reconsolidation rather than immediately reversing prior learning, an additional group of subjects (group 8) were retested a second time directly after the interference training

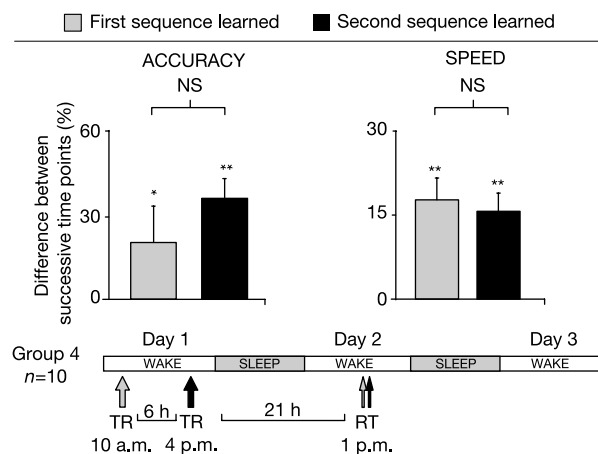


Figure 3 Changes in motor learning in group 4. Subjects initially learned the first motor sequence at 10 a.m. on day 1, and following a 6-hour waking interval, learned the second motor sequence. In contrast to group 2 who learned the second sequence immediately after the first, group 4 revealed significant improvements in performance speed ($P < 0.005$) and accuracy ($P < 0.05$) for both sequences.

on day 2, rather than on day 3 as in group 7, and showed no decrease in either speed or accuracy compared to the earlier retest (Fig. 4d).

Taken together with previous findings^{7,8,16}, these results describe a highly complex process of post-training memory consolidation and reconsolidation: (1) during initial training, learning occurs, leading to significant performance improvements; (2) between 10 minutes and 6 hours later, without intervening sleep, the memory undergoes the first, stabilization phase of consolidation, making it resistant to interference from a competing memory, but producing no improvement in speed or accuracy; (3) during the subsequent night, a second, enhancement stage of consolidation occurs, increasing both performance speed and accuracy on the task; (4) despite now having undergone both stabilization and enhancement, brief periods of rehearsal return the memory to a labile state, rendering it once again vulnerable to interference from a competing motor pattern, and in need of reconsolidation.

These findings provide, to our knowledge, the first clear dissociation of memory consolidation into temporally distinct phases of wake-state stabilization and sleep-dependent enhancement. However, it is unlikely that these stages of memory formation are generalizable across all human memory domains, as time-dependent consolidation has not been found for all memory tasks¹⁷. It also remains to be determined if these unique forms of memory consolidation are the result of common or distinct neural mechanisms. But the contrasting neurochemical and neurophysiological characteristics found across wake-sleep brain states¹⁸, together with their remarkably different patterns of gene expression and protein synthesis¹⁹, would argue for uniquely different neural mechanisms.

Our findings of memory lability and reconsolidation in humans complement similar findings in animal and clinical studies^{10–15}. Yet in these earlier reports, interference was produced using either electroconvulsive shock or cerebral injection of protein synthesis inhibitors, leaving it unclear whether more natural interventions could also block reconsolidation of labilized memories in humans. Our findings demonstrate that the simple process of training on a new motor sequence can block the reconsolidation of similar memories that have been returned to a labile state by simple rehearsal, and strongly imply functional significance for this process.

Allowing motor memories to return to a labile state requiring subsequent reconsolidation may permit the continued refinement and reshaping of previously learned movement skills in the context

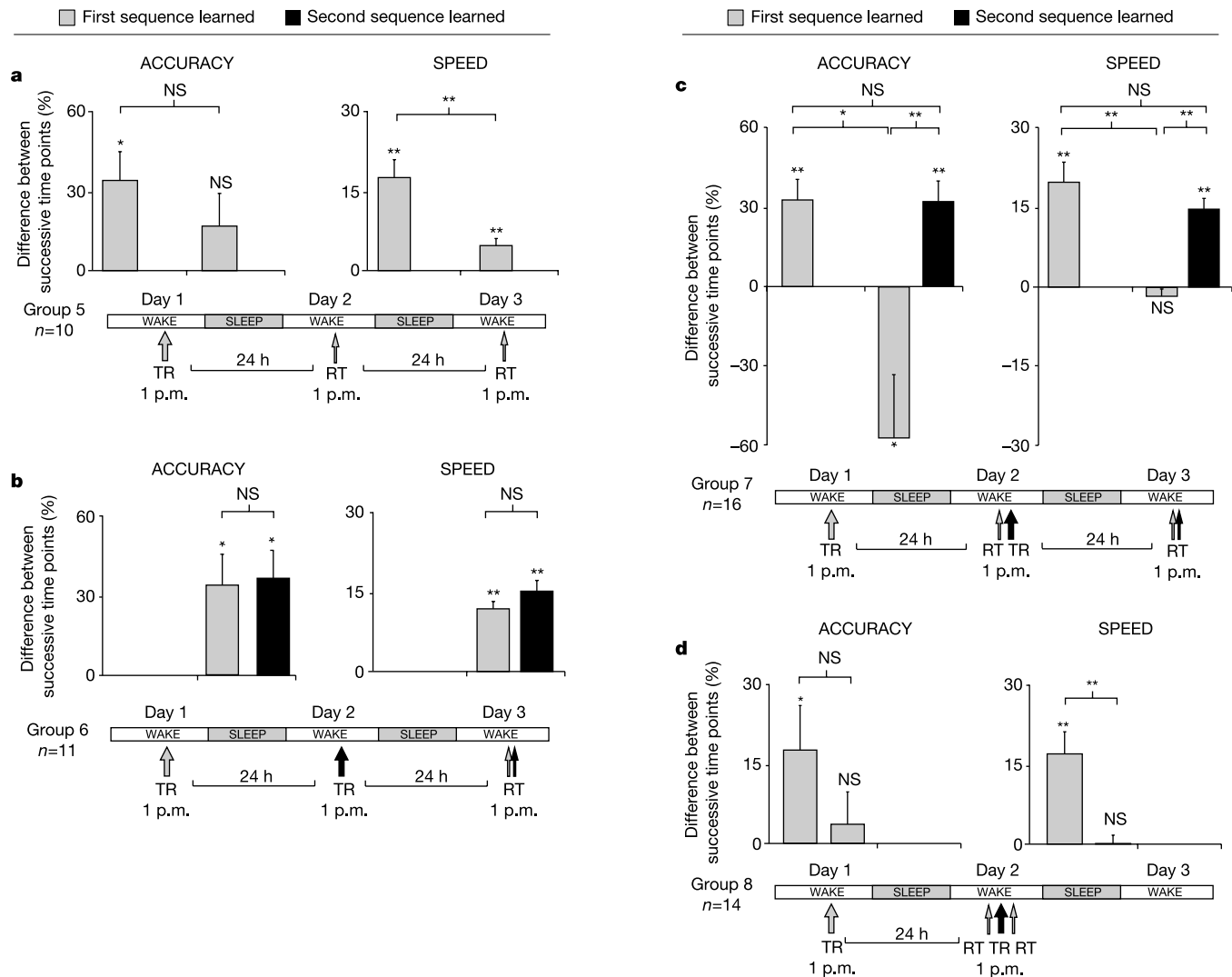


Figure 4 Changes in motor learning for groups 5–8. **a**, Group 5: subjects trained on only one sequence on day 1, and were retested once on day 2 and once on day 3. Significant improvements in accuracy ($P = 0.014$) and speed ($P < 0.001$) developed on day 2, with smaller but continued increases in speed ($P < 0.001$) and non-significant improvements in accuracy developing on day 3. **b**, Group 6: subjects were trained on the first sequence on day 1, and 24 hours later, trained on the second sequence on day 2. When retested on day 3, improvements in accuracy ($P < 0.015$) and speed ($P < 0.003$) had developed for both sequences equally. **c**, Group 7: subjects trained on the first sequence on day 1. On day 2, subjects were retested on the first sequence, demonstrating overnight improvements in speed ($P < 0.001$) and accuracy ($P < 0.001$) at the first retest. Immediately following this recall test on day 2, subjects learned the second sequence. When retested on this second sequence, overnight improvements in speed ($P < 0.001$)

and accuracy ($P < 0.001$) had developed on day 3, similar to those that developed for the first sequence on day 2. However, when subjects were now retested again on the first sequence on day 3 (+48 hours), the initial learning achieved on day 2 was lost, with accuracy decreasing significantly relative to day 2 ($P < 0.030$), and speed showing a non-significant reduction. **d**, Group 8: subjects trained on the first sequence on day 1 and were retested on day 2, showing significant overnight improvements in speed ($P = 0.001$) and accuracy ($P = 0.056$). Following recall of the first sequence on day 2, subjects learned the second interference sequence, but in contrast to group 7, were immediately retested on the first sequence again. At this second retest on day 2, subjects demonstrated maintenance, not loss, of the overnight improvements relative to the first retest (speed, $P = 0.947$; accuracy, $P = 0.447$).

of ongoing experience, and may even act as a process of extracting and integrating common elements from related complex movements. These results raise the possibility that similar mechanisms may also contribute to the integration of episodic memories and the revision of semantic knowledge based on newly acquired information. □

Received 3 June; accepted 16 July 2003; doi:10.1038/nature01930.

- McGaugh, J. L. Memory—a century of consolidation. *Science* **287**, 248–251 (2000).
- Mueller, W. et al. Early consolidation in human primary motor cortex. *Nature* **415**, 640–644 (2002).
- Brashers-Krug, T., Shadmehr, R. & Bizzi, E. Consolidation in human motor memory. *Nature* **382**, 252–255 (1996).

- Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. & Sagi, D. Dependence on REM sleep of overnight improvement of a perceptual skill. *Science* **265**, 679–682 (1994).
- Smith, C. & MacNeill, C. Impaired motor memory for a pursuit rotor task following Stage 2 sleep loss in college students. *J. Sleep Res.* **3**, 206–213 (1994).
- Stickgold, R., James, L. & Hobson, J. A. Visual discrimination learning requires sleep after training. *Nature Neurosci.* **3**, 1237–1238 (2000).
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A. & Stickgold, R. Practice with sleep makes perfect: Sleep dependent motor skill learning. *Neuron* **35**, 205–211 (2002).
- Walker, M. P., Brakefield, T., Hobson, J. A. & Stickgold, R. Sleep and the time course of motor skill learning. *Learn. Mem.* **10**, 275–284 (2003).
- Nader, K. Memory traces unbound. *Trends Neurosci.* **26**, 65–72 (2003).
- Misanin, J. R., Miller, R. R. & Lewis, D. J. Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science* **160**, 554–555 (1968).
- Rubin, R. D. Clinical use of retrograde amnesia produced by electroconvulsive shock. A conditioning

- hypothesis. *Can. Psychiatr. Assoc. J.* **21**, 87–90 (1976).
12. Gordon, W. C. & Spear, N. E. The effects of strychnine on recently acquired and reactivated passive avoidance memories. *Physiol. Behav.* **10**, 1071–1075 (1973).
 13. Nader, K., Schafe, G. E. & Le Douarin, J. E. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* **406**, 722–726 (2000).
 14. Debiec, J., LeDoux, J. E. & Nader, K. Cellular and systems reconsolidation in the hippocampus. *Neuron* **36**, 527–538 (2002).
 15. Milekic, M. H. & Alberini, C. M. Temporally graded requirement for protein synthesis following memory reactivation. *Neuron* **36**, 521–525 (2002).
 16. Fischer, S., Hallschmid, M., Elsner, A. L. & Born, J. Sleep forms memory for finger skills. *Proc. Natl Acad. Sci. USA* **99**, 11987–11991 (2002).
 17. Goedert, K. M. & Willingham, D. B. Patterns of interference in sequence learning and prism adaptation inconsistent with the consolidation hypothesis. *Learn. Mem.* **9**, 279–292 (2002).
 18. Pace-Schott, E. F. & Hobson, J. A. The neurobiology of sleep: Genetics, cellular physiology and subcortical networks. *Nature Rev. Neurosci.* **3**, 591–605 (2002).
 19. Hobson, J. A. & Pace-Schott, E. F. The cognitive neuroscience of sleep: Neuronal systems, consciousness and learning. *Nature Rev. Neurosci.* **3**, 679–693 (2002).

Acknowledgements We thank B. Kocsis, E. Pace-Schott and R. Fosse for comments regarding the study. This research was supported by the National Science Foundation and the National Institute of Mental Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.P.W. (mwalker@hms.harvard.edu).

Environmentally mediated synergy between perception and behaviour in mobile robots

Paul F. M. J. Verschure, Thomas Voegtlin* & Rodney J. Douglas

Institute of Neuroinformatics, University/Swiss Federal Institute of Technology (ETH) Zürich, CH-8057 Zürich, Switzerland

* Present address: Institute for Theoretical Biology, Humboldt-University Berlin, D-10115 Berlin, Germany

The notion that behaviour influences perception seems self-evident, but the mechanism of their interaction is not known. Perception and behaviour are usually considered to be separate processes. In this view, perceptual learning constructs compact representations of sensory events, reflecting their statistical properties^{1,2}, independently of behavioural relevance^{3,4}. Behavioural learning^{5,6}, however, forms associations between perception and action, organized by reinforcement^{7,8}, without regard for the construction of perception. It is generally assumed that the interaction between these two processes is internal to the agent, and can be explained solely in terms of the neuronal substrate⁹. Here we show, instead, that perception and behaviour can interact synergistically via the environment. Using simulated and real mobile robots, we demonstrate that perceptual learning directly supports behavioural learning and so promotes a progressive structuring of behaviour. This structuring leads to a systematic bias in input sampling, which directly affects the organization of the perceptual system. This external, environmentally mediated feedback matches the perceptual system to the emerging behavioural structure, so that the behaviour is stabilized.

One reason for the lack of progress in understanding the interrelationship of behaviour and perception is experimental intractability. An explanation of their coupling requires detailed analysis at both the behavioural and neuronal levels. Our approach was to bypass the animal experimental difficulty by using a mobile robot, for which it is possible to fully observe and quantify perception and

behaviour. The robot is controlled by a neural model, called distributed adaptive control (DAC), that includes mechanisms for perceptual and behavioural learning^{10,11}. The DAC architecture (see Fig. 1 and Methods) consists of three layers: 'reactive', 'adaptive' and 'contextual' control.

The reactive control layer implements a repertoire of basic reflex actions where low-complexity sensory events, unconditioned stimuli (US), trigger simple actions, unconditioned responses (UR), via an internal state (IS) representation. As a result of learning at the level of adaptive control, the purely reactive activation of the IS populations by US events is progressively replaced by acquired representations of sensory events, conditioned stimuli (CS), and the generation of conditioned responses (CR)^{11,12}. US events are the initial reinforcers of this learning process. The local learning mechanism that is used automatically generates a measure, D , of the discrepancy between expected and actual CS events (see Methods). When D falls below a specified transition threshold, θ_D , the contextual control layer is enabled. This layer is a behavioural learning system that constructs higher-order representations of the temporal order of the sensori-motor representations constructed by the adaptive layer (see Methods).

Representations of CS and CR events are stored in short-term memory (STM) when the adaptive layer triggers CRs. The content of STM is stored in long-term memory (LTM) when a goal state is reached, such as when a target is found. CS representations of the LTM of the contextual layer are matched to those generated by the adaptive layer. The best-matching CS representation at the level of contextual control will define the next action by projecting its CR representation onto the motor population M when the reactive layer is quiescent. Chaining through a LTM sequence is achieved through a biased competition mechanism (see Methods). DAC is a practical model of how different learning systems in the mammalian brain act together to generate adaptive goal-oriented behaviour, and is a standard in the field of new artificial intelligence and behaviour-based robotics^{13–16}. Moreover, it exhibits the regularities of bayesian decision-making that are thought to be one of the characteristics of human cognition^{17,18}.

We first investigated the hypothesis that the performance of the robot is enhanced through the contextual layer. To test this hypothesis, we used both simulated and real-world robots in a foraging task where collisions had to be minimized while the number of targets found had to be maximized. We distinguished

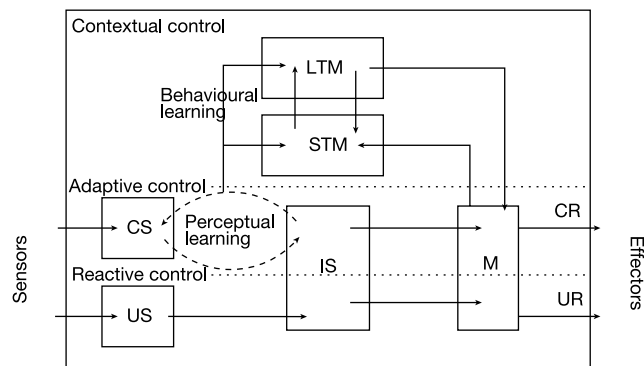


Figure 1 Distributed adaptive control. DAC is based on the assumption that adaptive behaviour results from three tightly coupled layers of control: reactive, adaptive and contextual control. Each box represents a neuronal population. Arrows indicate the connections between these populations. US, unconditioned stimulus population. CS, conditioned stimulus population. IS, internal state populations. M, motor neuron population. UR, unconditioned response. CR, conditioned response. STM, short-term memory. LTM, long-term memory. See text and Methods for explanation.

two conditions, in which the contextual layer was either ‘enabled’ or ‘disabled’.

In simulation experiments (see Fig. 2a and Supplementary Information), we found that the adaptive layer improved the performance of the robot through a learning-dependent avoidance of collisions, reflected in the increase of the target/collision ratio (Fig. 2b). We also observed that at the onset of the second stimulation period, the performance of the two conditions diverges: in the enabled condition, performance is strongly enhanced compared with the disabled condition. This difference is due to the activation of the contextual control layer in the enabled condition, as can be deduced from the evolution of the discrepancy measure D (Fig. 2c)—shortly after the onset of the second stimulation period, the D value falls below the transition threshold θ_D .

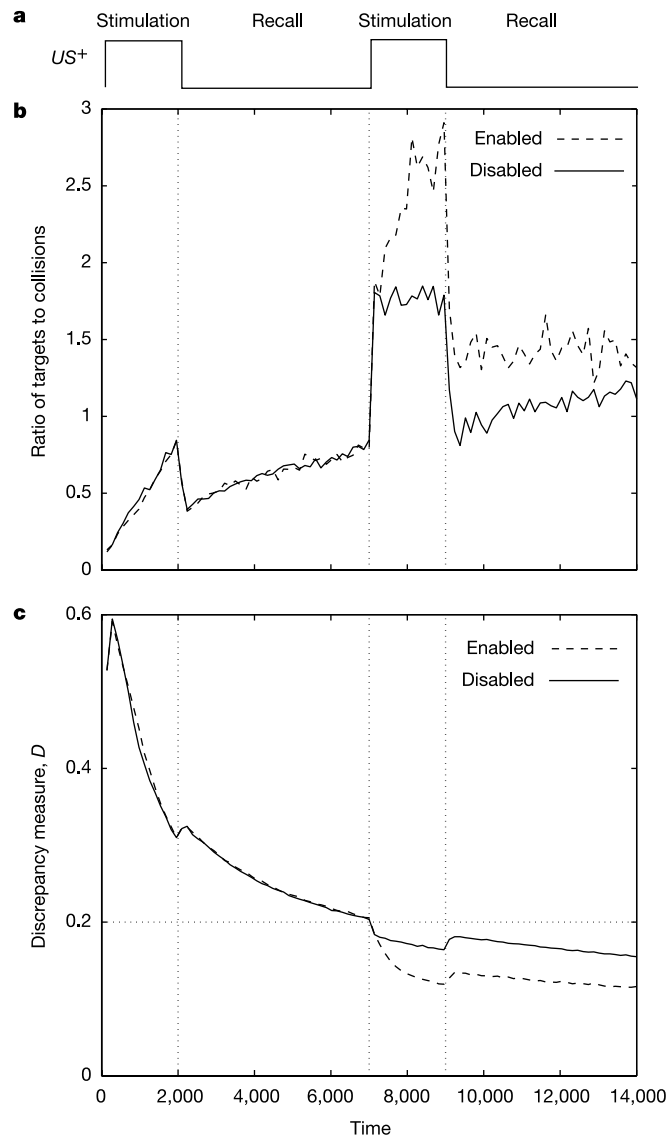


Figure 2 Experimental protocol and performance in simulated robot experiments for the disabled and enabled conditions using 1,000 exemplars per condition. The initial position and orientation of the robots were randomized. **a**, Each experiment consists of two cycles. Each cycle commences with a stimulation period (2,000 time steps), in which the targets emit a signal (US^+), followed by a recall period (5,000 time steps), in which the target signals are absent. **b**, The ratio between the number of targets found and accumulated collisions averaged in non-overlapping time windows of 100 time steps. **c**, Evolution of the discrepancy measured, D . The dotted horizontal line indicates the value of the transition threshold, θ_D .

During the second stimulation and recall periods, the D value of the enabled condition is markedly below that of the disabled condition. This reduction is accompanied by a significantly lower value of the average absolute change in synaptic efficacies of the connections between the CS and IS populations: enabled $\|\Delta W\| = 2.4 \times 10^{-2}$, disabled $\|\Delta W\| = 2.6 \times 10^{-2}$ (t -test, $P \ll 0.001$). Hence, the transition to contextual control leads to a reduction of the discrepancy between predicted and actual CS events and to a stabilization of the synaptic weights of the adaptive control layer. However, our model has no internal feedback from the contextual to the adaptive control layer: D and ΔW are properties local to the perceptual learning system. Therefore, this difference must be due to the difference in the overt behaviour generated in the two conditions and the systematic bias in the sampling of CS events that this difference causes, that is, behavioural feedback. That is, behaviour is less variable when the contextual layer is enabled, thereby reducing the variability of the sampled sensory inputs. We tested this hypothesis by comparing the entropies of behaviour and sampled stimuli between the two conditions (see Supplementary Information).

We characterized the behavioural entropy, H_B , of the distribution of positions visited for both conditions in an experiment of 10^6 time steps, using the same protocol as above (Fig. 2a). In the disabled condition, H_B was 15.1, whereas the enabled condition showed a lower H_B value of 14.2. These numbers can be compared with the maximal entropy of 15.4, obtained from a uniform distribution of positions; and to an H_B of 11.2 for a minimal cyclic trajectory that follows the shortest path between subsequent targets. The difference in H_B between a uniform distribution of positions and the disabled condition can be explained by the learned avoidance behaviour that causes the robot to avoid the regions close to obstacles. The additional reduction of H_B in the enabled condition is due to a further task-dependent structuring of behaviour (see also real robot results below).

We assessed whether the structuring of behaviour quantified by H_B is associated with a change in the input statistics by calculating the sampling entropy, H_S , of the states of the CS-related sensor. H_S was lower (6.80) for the enabled condition than for disabled (7.95), demonstrating that the structuring of behaviour displayed in the enabled condition is associated with a marked reduction in the variability of the sampled CS events. We deduce that the reduction in behavioural variability, which results from behavioural learning, reduces the set of inputs that the perceptual learning system must classify. As a result of this behavioural feedback, the structures for perceptual learning become adjusted to a smaller set of input states. This change is reflected in the reduction and stabilization of D (Fig. 2c) and the reduced value of the synaptic changes we observed in the enabled condition compared with disabled.

So far we have demonstrated the effect of behavioural feedback on perceptual learning. However, its implications extend beyond sensory classification alone. In our model, the transition to the use of contextual control occurs when D falls below a fixed threshold, θ_D . Hence, behavioural feedback, because it reduces D , may favour the transition to contextual control. To test this hypothesis, we recorded the downward crossings through θ_D for the experiments reported in Fig. 2. On average, these transitions occurred near the onset of the second stimulation period (Fig. 2c). However, in individual experiments, D does not decrease monotonically and small fluctuations of D around θ_D , in the enabled condition, result in oscillations between activating or deactivating contextual control. For the disabled condition, we observed that, on average, oscillations around the transition threshold occurred 6.31 times, whereas for the enabled condition these oscillations were strongly reduced to 3.86. Because there is no mechanism in our model that stabilizes the transition to contextual control, we conclude that this stabilization occurs through behavioural feedback. The switch to contextual control indirectly reduces D through behaviour, and this

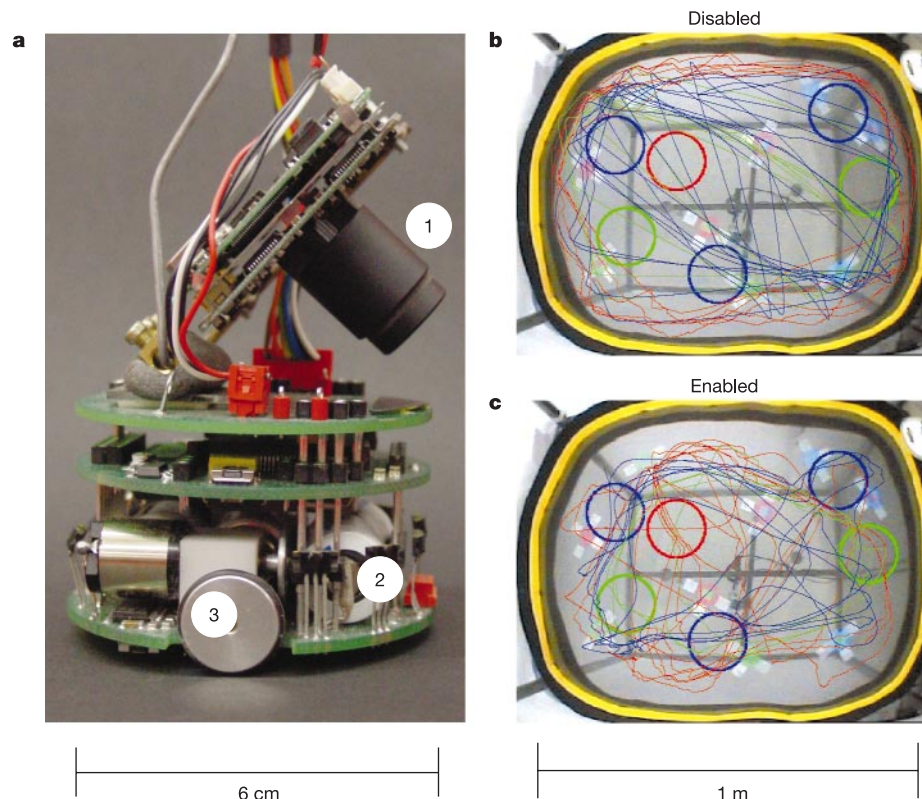


Figure 3 Micro-robot experiments. **a**, The CS is provided by a colour charge-coupled device (CCD) camera (1), whereas aversive and appetitive USs, defined as collisions and the level of ambient light, respectively, are provided by the six frontal infrared sensors (2). Locomotion is provided by two wheels (3). **b, c**, The environment is surrounded by a yellow

wall and coloured patches are placed on the floor. The targets (coloured circles) correspond to light sources. The trajectories displayed are those generated by the robot during the recall period, for conditions disabled (**b**) and enabled (**c**). Red, one light source. Green, two light sources. Blue, three light sources.

reduction of D in turn directly favours contextual control. Hence, our simulation experiments show that behavioural feedback not only matches the perceptual system to the emerging behavioural structure, but also in turn stabilizes behavioural learning.

We assessed the generality of our results by applying our model to a similar foraging task using the real-world robot Khepera¹⁹ (Fig. 3a and Supplementary Information). The task requires the robot to learn to avoid collisions (US^-) while visiting illuminated regions in the environment (US^+) as often as possible. The CSs are defined by different colours present in the environment, that is, a yellow wall and coloured patches on the floor.

The trajectories during the recall test show an organization of behaviour consistent with the behavioural entropy measures obtained in the simulation experiments (Fig. 3b, c). In the disabled condition (Fig. 3b), the robot associated yellow patches with aversive events and learned to avoid collisions with the walls of the arena. However, the overall behaviour of the robot consists of following the wall interspersed with deflections off the wall. Thus, the target areas are visited only by chance. In the enabled condition (Fig. 3c), the robot used the coloured patches to reliably locate the areas where it had found positive reinforcement in the past, leading to a more structured behaviour. We quantified this structure by recasting the behaviour as a Markov process. We used an environment with two targets (see Fig. 3b, c, green circles) and estimated the probabilities of the transition between pairs of CS events excluding the yellow wall (Fig. 4 and Supplementary Information).

The disabled condition shows probabilities below 0.25 in transitions between 10 CS pairs. Most of these transitions occur in the periphery of the environment. In the enabled condition, the maximum transition probability increases to 0.75 for a total of 20

CS pairs, including colour patches in the centre of the environment. This pattern is consistent with the trajectories observed earlier (Fig. 3b, c). Moreover, the variability of $\|\Delta W\|$ was much smaller in the enabled compared with the disabled condition consistent with the values found in the simulation experiments: enabled $\|\Delta W\| = 3 \times 10^{-3}$, disabled $\|\Delta W\| = 8 \times 10^{-2}$. This confirms that our results fully generalize to the real world and are robust with respect to the details of the sensory and motor systems employed.

To directly evaluate the effect of behavioural feedback on performance, we compared the enabled condition with a separate control condition, called 'static'. In the static condition, the synaptic efficacies of the adaptive layer were fixed and initialized with the values of those of the enabled condition when the latter stored its first LTM sequence. The static condition can still store STM and LTM sequences but its perceptual learning system is switched off. We found, using both simulated and real-world robots, that performance in the static condition was strongly reduced by comparison with the enabled condition. Moreover, its behavioural trajectory was more variable, its D value higher and the oscillations around θ_D enhanced (see Supplementary Information for details). This result confirms that behavioural feedback directly enhances performance.

Although the microscopic and local view of behaviour and perception has been questioned for being too restricted^{20,21}, it is unclear which other factors should be included to comprehend fully how they are shaped through experience. The robot experiments reported here demonstrate that learning-dependent changes in behaviour can establish a macroscopic feedback loop. Once the contextual layer is activated, the robot can retain and recall

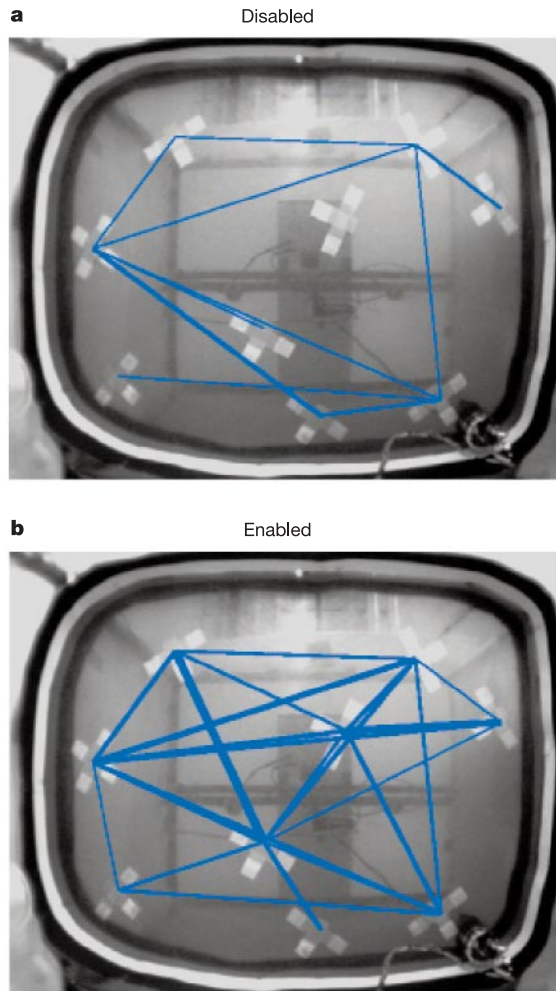


Figure 4 Quantification of the emerging behavioural structure using a hidden Markov model. The transition probabilities between pairs of CS events are mapped on the environment for conditions disabled (**a**) and enabled (**b**). Probabilities exceeding 0.1 are represented by a line that connects the corresponding colour patches. Line thickness reflects the value of the transition probability. Transitions that include the yellow surrounding wall are omitted. The maximum value of the transition probability is 0.25 for the disabled condition and 0.75 for the enabled condition.

successful trajectories, and it develops a habit of following mainly these established paths. In this more restricted and therefore less variable environment, the task of the perceptual learning system is facilitated and biased towards behaviourally relevant sensory events. This improved sensory classification in turn stabilizes the emerging behavioural patterns acquired by the behavioural learning system. Hence, non-neuronal feedback can directly contribute to the organization of behaviour and perception, establishing a synergistic interaction.

To what extent are our robotic results relevant to the understanding of the brain? One prediction of our study is that a change in the afferent input to sensory areas, due to behavioural feedback, can systematically change the organization of perceptual systems. This prediction is supported by the observation that, during development, the organization and response properties of primary sensory areas can be strongly influenced by their afferent inputs²², indicating that during these stages of development behavioural feedback can have a direct impact on the organization of perceptual systems. We expect that behavioural feedback would affect perceptual systems and those structures that readout from these systems,

both during early development and beyond. The observation that the place fields of the CA1 region of the hippocampus show an experience-dependent expansion and shift in their centre of mass in relation to the behavioural trajectory supports this suggestion²³. Moreover, the observation that the responses in the primary auditory cortex are attenuated during vocalization suggests that the brain actively regulates the impact of behavioural feedback²⁴. Thus, we propose that the brain exploits behavioural feedback to constrain perceptual learning and to stabilize acquired behavioural structures.

The role of feedback internal to the neuronal systems underlying perceptual and behavioural learning is receiving increasing attention, particularly in the context of prediction errors of stimulus properties and rewards^{25,26}. Behavioural feedback modifies stimulus sampling and so provides an additional extra-neuronal path for the reduction of prediction errors. Our results provide direct evidence for this assertion by demonstrating that the environmentally mediated feedback between behaviour and perception can significantly affect both of these processes. □

Methods

All neural elements of the DAC architecture are implemented as linear threshold units.

Reactive control

The reactive control layer implements a set of pre-wired reflexes that support basic behaviours where low-complexity sensory events, the US, trigger immediate actions, the UR. The relationship between the US and the UR is established through an intermediate stage that reflects an IS, for example, aversion, ($US^- \rightarrow IS^-$) or appetite ($US^+ \rightarrow IS^+$). US events activate neurons in IS, which in turn activate a population of motor neurons (M), producing an avoidance or approach action (UR), respectively. Action selection is defined by a winner-take-all mechanism in neuronal population M . Conflicts between approach and avoidance actions are resolved by pre-wired relationships between IS populations. If none of the IS populations is active, the reactive control layer generates exploration behaviour, consisting of translational movements.

Adaptive control

The adaptive control layer provides mechanisms for perceptual learning and constructs representations of complex sensory events (CS). These representations arise from the experience-dependent changes in the efficacies of the synaptic connections between the populations of neurons reflecting sensory events (CS) and IS. Activity in IS populations, A_{IS} , is defined by:

$$A_{IS} = A_{US} + W A_{CS} \quad (1)$$

where A_{US} and A_{CS} denote the activity of the US-conveying sensor and CS-driven neural activity, respectively. W represents the synaptic efficacies of the connections between CS and IS populations. The change of these synaptic efficacies, ΔW , is defined by:

$$\Delta W = \eta A_{IS} (A_{CS} - \gamma W^T A_{IS}) \quad (2)$$

where η is a learning rate, γ a linear gain and W^T the transpose of W . Hence, ΔW depends on the difference, or reconstruction error, between the actual CS state, A_{CS} , and that predicted CS state, given the current state of IS and W , $W^T A_{IS}$, also referred to as the sensory expectation, E . This learning rule is related to filtering theory and state estimation²⁷, and similar approaches have been applied to modelling the cortical mechanisms of perceptual learning². Recently, a biophysically realistic implementation of this predictive learning rule was presented²⁸.

D is defined as a leaky average of the difference, $d(A_{CS}, E)$, between the current, A_{CS} , and the predicted CS state, E :

$$D(t+1) = (1 - \alpha_D) D(t) + \alpha_D d(A_{CS}, E)(t) \quad (3)$$

where α_D defines the integration time constant and $d(A_{CS}, E)$ is defined as:

$$d(A_{CS}, E)(t) = \frac{1}{N} \sum_{j=1}^N \left| \frac{A_{CS}(j)}{\max_{1 \leq j' \leq N} A_{CS}(j')} - \frac{E(j)}{\max_{1 \leq j' \leq N} E(j')} \right| \quad (4)$$

where N stands for the size of the CS population.

Contextual control

The contextual control layer provides mechanisms of STM and LTM and is enabled once the discrepancy measure D falls below a fixed threshold, θ_D . This ensures that the representations of CS events, E , used in STM and LTM are based on stable classifications constructed by the perceptual learning system of the adaptive layer. Some key aspects of the contextual layer are: (1) salient CS events, E , and their associated actions, M , are stored in STM, conserving their order of occurrence. Salience is defined as the interruption of exploration behaviour in the absence of US events. (2) The content of STM is stored in LTM when a goal state is reached, such as finding a target. (3) The sensory content of LTM segments is continuously matched against interpreted sensory events, E , generated by the adaptive layer. (4) LTM segments that match E compete for behavioural control. (5) If the best-matching segment exceeds a specific matching threshold it will control behaviour,

provided that the reactive control layer is inactive. (6) Chaining through LTM sequences is achieved by biasing the LTM matching process.

The competition among LTM segments takes place on a quantity m where m_l^k of segment l of sequence k is defined as:

$$m_l^k = c_l^k t_l^k \quad (5)$$

where c_l^k is defined as $d(E, E_l^k)$ using the definition of d provided in equation (4). t_l^k , $t_l^k \in [0; 1]$, is a dynamic threshold that provides a probabilistic mechanism for chaining through a LTM sequence. In case segment l of sequence k wins the competition by virtue of having the lowest value of m , it will reduce t of segment $l + 1$, of k , t_{l+1}^k , to a fixed value β , $\beta \in [0; 1]$. t_l^k relaxes to its default value of 1 according to $t_l^k(t + 1) = \alpha_T + (1 - \alpha_T)t_l^k(t)$, $\alpha_T \in [0; 1]$. The LTM segment that wins the competition will dominate the behavioural output of the overall system when its m is below a fixed threshold.

In the present implementation, STM is a ring buffer with a fixed length of 25 segments, while the capacity of LTM is limited to 64 sequences.

Received 3 March; accepted 7 August 2003; doi:10.1038/nature02024.

- Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
- Rao, R. & Ballard, D. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neurosci.* **2**, 79–87 (1999).
- Logothetis, N. & Sheinberg, D. Visual object recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
- Goldstone, R. Perceptual learning. *Annu. Rev. Psychol.* **49**, 585–612 (1998).
- Mackintosh, N. *The Psychology of Animal Learning* (Academic, New York, 1974).
- Lavond, D. G., Kim, J. J. & Thompson, R. F. Mammalian brain substrates of aversive classical conditioning. *Annu. Rev. Psychol.* **44**, 317–342 (1993).
- Thorndike, E. Animal intelligence: an experimental study of the associative processes in animals. *Psychol. Rev. Ser. Monogr. Suppl.* **2**, 1–109 (1898).
- Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, Massachusetts, 1998).
- Squire, L. & Kandel, E. *Memory: From Mind to Molecules* (Scientific American Library, New York, 1999).
- Verschure, P. F. M. J., Kröse, B. & Pfeifer, R. Distributed adaptive control: The self-organization of structured behavior. *Rob. Auton. Syst.* **9**, 181–196 (1992).
- Verschure, P. F. M. J. & Voegtlin, T. A bottom-up approach towards the acquisition, retention, and expression of sequential representations: Distributed adaptive control III. *Neural Netw.* **11**, 1531–1549 (1998).
- Verschure, P. F. M. J. & Pfeifer, R. in *From Animals to Animats: Proc. 2nd Int. Conf. Simul. Adapt. Behav. (Honolulu, Hawaii)* (eds Meyer, J. A., Roitblat, H. & Wilson, S.) 210–217 (MIT Press, Cambridge, Massachusetts, 1992).
- McFarland, D. & Bosser, T. *Intelligent Behavior in Animals and Robots* (MIT Press, Cambridge, Massachusetts, 1993).
- Clancey, W. *Situated Cognition: On Human Knowledge and Computer Representations* (Cambridge University Press, Cambridge, UK, 1996).
- Arkin, R. *Behavior-Based Robotics* (MIT Press, Cambridge, Massachusetts, 1998).
- Pfeifer, R. & Scheier, C. *Understanding Intelligence* (MIT Press, Cambridge, Massachusetts, 1999).
- Verschure, P. F. M. J. & Althaus, P. A real-world rational agent: Unifying old and new AI. *Cogn. Sci.* **27**, 561–590 (2003).
- Massaro, D. *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, Massachusetts, 1997).
- Mondada, F., Franzi, E. & Jenne, P. *Experimental Robotics III: Proc. 3rd Int. Symp. Exp. Rob. (Kyoto, Japan, 28–30 October 1993)* 501–513 (Springer, Berlin, 1993).
- Tolman, E. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).
- Bell, A. Levels and loops: the future of artificial intelligence and neuroscience. *Phil. Trans. R. Soc. Lond. B* **354**, 2013–2020 (1999).
- Sur, M. & Leamy, C. Development and plasticity of cortical areas and networks. *Nature Rev. Neurosci.* **2**, 251–261 (2001).
- Mehta, M., Barnes, C. & McNaughton, B. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl Acad. Sci. USA* **94**, 8918–8921 (1997).
- Houde, J., Nagarajan, S., Sekihara, K. & Merzenich, M. Modulation of the auditory cortex during speech: An MEG study. *J. Cogn. Neurosci.* **14**, 1125–1138 (2002).
- Rescorla, R. & Wagner, A. in *Classical Conditioning 2. Current Theory and Research* (eds Black, A. H. & Prokasy, W. E.) 64–99 (Appleton-Century-Crofts, New York, 1972).
- Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).
- Kalman, R. A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* **82**, 35–45 (1960).
- Sanchez-Montanes, M., Verschure, P. F. M. J. & König, P. Local and global gating of plasticity. *Neural Comput.* **12**, 519–529 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Baumgartner and J. Manzolli for their support in performing the Markov analysis. Part of this research is supported by the Swiss National Science Foundation, the Volkswagen foundation and the Körber Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.F.M.J.V. (pfmjv@ini.phys.ethz.ch).

Regulation of neuroblast competence in *Drosophila*

Bret J. Pearson & Chris Q. Doe

Institutes of Neuroscience and Molecular Biology, Howard Hughes Medical Institute, University of Oregon 1254, Eugene, Oregon 97403, USA

Individual neural progenitors generate different cell types in a reproducible order in the retina^{1–3}, cerebral cortex^{4–6} and probably in the spinal cord⁷. It is unknown how neural progenitors change over time to generate different cell types. It has been proposed that progenitors undergo progressive restriction⁸ or transit through distinct competence states^{9,10}; however, the underlying molecular mechanisms remain unclear. Here we investigate neural progenitor competence and temporal identity using an *in vivo* genetic system—*Drosophila* neuroblasts—where the Hunchback transcription factor is necessary and sufficient to specify early-born cell types¹¹. We show that neuroblasts gradually lose competence to generate early-born fates in response to Hunchback, similar to progressive restriction models⁸, and that competence to acquire early-born fates is present in mitotic precursors but is lost in post-mitotic neurons. These results match those observed in vertebrate systems, and establish *Drosophila* neuroblasts as a model system for the molecular genetic analysis of neural progenitor competence and plasticity.

Despite substantial progress in vertebrates, we still know little about the molecular basis for how a single neural precursor sequentially generates different cell types. This is primarily due to the lack of an *in vivo* model system where a single neural progenitor can be studied at reproducible times during its lineage. The *Drosophila* embryonic central nervous system (CNS) lends itself well to the study of neural progenitor plasticity because neural progenitors (neuroblasts (NBs)) can be individually identified, each NB generates different cell types in a reproducible order, molecular markers exist for each of these cell types, intrinsic factors are known that confer different temporal identities, and gene expression can be readily manipulated at specific points within the NB lineage¹¹. NBs repeatedly divide in a stem-cell-like mode to ‘bud off’ a series of smaller daughter cells called ganglion mother cells (GMCs). Cell lineage studies show that every GMC has a unique identity based on its ‘birth’ order within the NB lineage, and generates a characteristic pair of neurons or glia. Recently, four transcription factors have been identified that are excellent candidates for specifying GMC temporal identity^{11,12}. NBs sequentially express the transcription factors Hunchback (Hb)→Krüppel→Pdm1→Castor, with GMCs inheriting the transcription-factor profile of the parental NB on their generation, which is then maintained in their own neuronal progeny (Fig. 1)¹¹. *hb* is both necessary and sufficient for specifying the first-born temporal identities in multiple NB lineages, even though first-born cells can be motor neurons, interneurons, or glia¹¹. Here we manipulate the timing and levels of Hb in a model NB lineage (NB7-1) to ask two fundamental questions: when does a NB lose competence to generate first-born neurons (immediately after Hb downregulation, progressively during its lineage, or never)? And when during neuronal differentiation is competence to generate first-born neurons lost (in NBs, GMCs, or post-mitotic neurons)?

To assay changes of temporal identity within a single lineage, we precisely define the birth order and sibling relationships of the early NB7-1 lineage. NB7-1 generates five motor neurons (U1–U5) and about 30 interneurons^{13,14}. The five U motor neurons have stereotyped positions in the CNS (Fig. 1a), express the Even-skipped (Eve) transcription factor (Fig. 2a), and innervate specific body wall

muscles^{11,13,15}. It has been proposed that U1–U5 develop from the first five GMCs in the lineage^{11,16}. Here we use a genetic method¹⁷ to induce permanent, positively marked cell clones at different points of the NB7-1 lineage. We find that the first five GMCs (GMC1–5) sequentially generate the U1–U5 motor neurons, and that each *Eve*⁺ motor neuron has an *Eve*[−] sibling (Fig. 1b; Supplementary Movie S1).

To test whether NBs lose competence to make first-born fates as they progress through their cell lineages, we use *prospero-Gal4* to re-express high levels of Hb in the NB7-1 lineage just after the birth of the Hb-negative GMC3 (Fig. 2b, cartoon). This has no effect on the normal U1–U3 cell fates, as expected, but subsequently the NB generates an average of 6.3 ($n = 60$) extra U1 motor neurons based on molecular marker expression (Fig. 2b and Table 1, row 2). To more rigorously determine whether these neurons have a U1 identity, we assay the axon projections of single or small groups of the putative U1 neurons, and observe that the ectopic U1 motor neurons project out of the intersegmental nerve to dorsal body wall

muscles, consistent with the normal U1 neuron (Fig. 2d; see also Supplementary Movie S2). We conclude that NB7-1 shows plasticity: first-born U1 motor neurons can be induced even after downregulation of endogenous Hb.

Hb is required to activate and repress target genes in a concentration-dependent manner along the anterior–posterior body axis during segmentation¹⁸. To test the idea that different levels of Hb might have different functions in the CNS, we repeated the previous experiment (Fig. 2b) but with lower levels of Hb expressed in precisely the same temporal pattern (Fig. 2c). Whereas high levels of Hb generate extra U1 neurons (Fig. 2b and Table 1, row 2), low levels of Hb produce presumptive U2 neurons (Fig. 2c and Table 1, row 3; we call them ‘presumptive’ U2 neurons because we cannot distinguish them from U3 motor neurons without using Hb as a marker). We conclude that high levels of Hb can induce U1 fates, whereas lower levels can induce presumptive U2 fates. This may explain how Hb generates two distinct temporal identities in the NB7-1 lineage (U1 and U2).

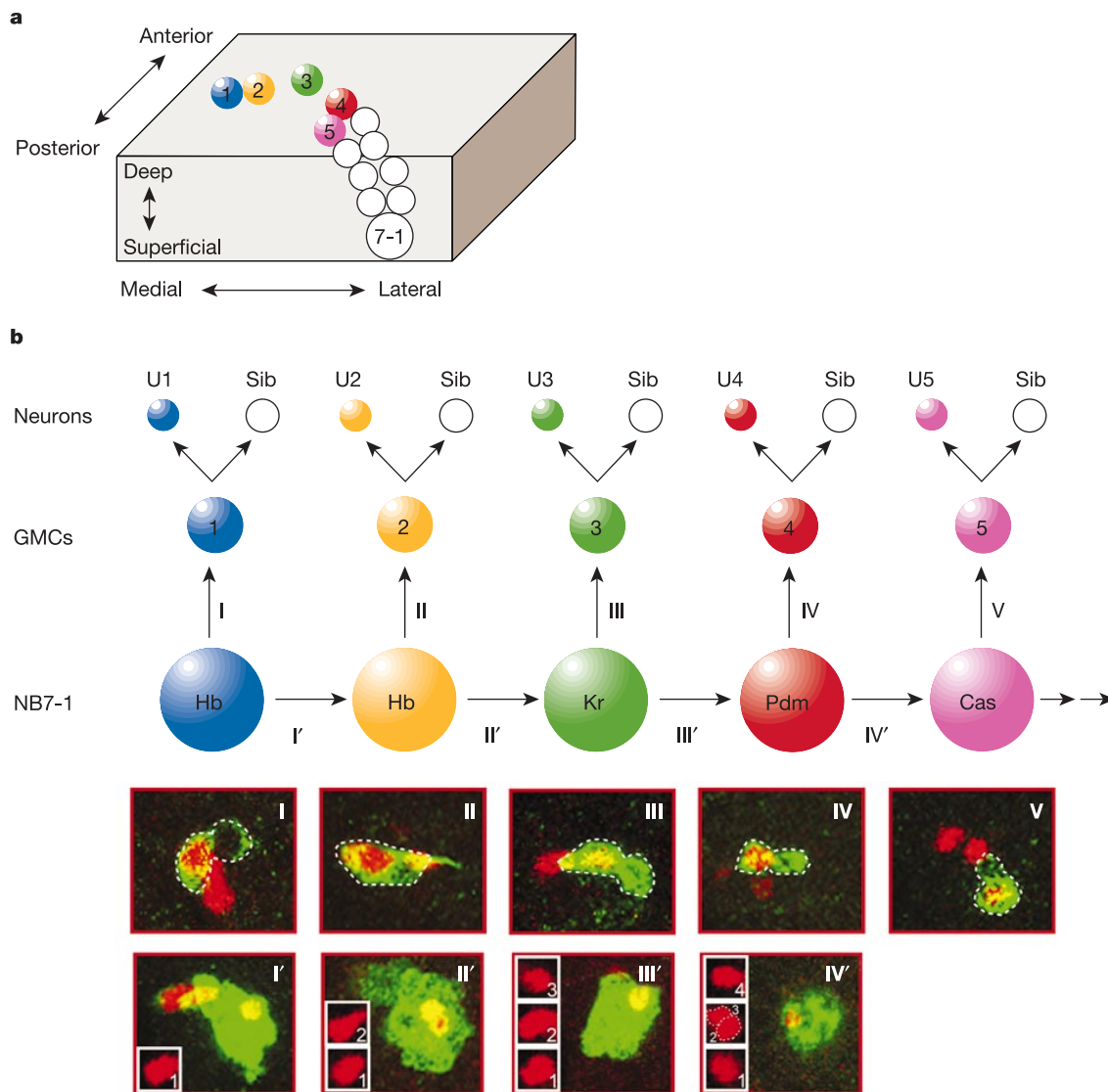


Figure 1 The early lineage of NB7-1. **a**, The *Eve*⁺ U1–U5 motor neurons (1–5) have stereotyped positions within the mature CNS¹⁴. One hemi-segment is shown.

b, Top: Schematic of the early NB7-1 lineage, based on data shown below. *Eve*⁺ U motor neuron, coloured; *Eve*[−] sibling cell, white. Bottom: Clonal analysis of the U1–U5 motor neurons (see Methods and Supplementary Movie 1 for details). Anterior, up; medial, left.

Clones induced in GMCs label two sibling neurons (*Tau*::βgal, green; clone outlined) that always contain one *Eve*⁺ U motor neuron (red) and one *Eve*[−] sibling cell (I–V). Clones induced in the NB label all progeny generated after the event (*Tau*::βgal, green; clone outlined) (I'–IV'). Insets show *Tau*::lacZ-negative U1, U2, U3 or U4 motor neurons from the same hemi-segment.

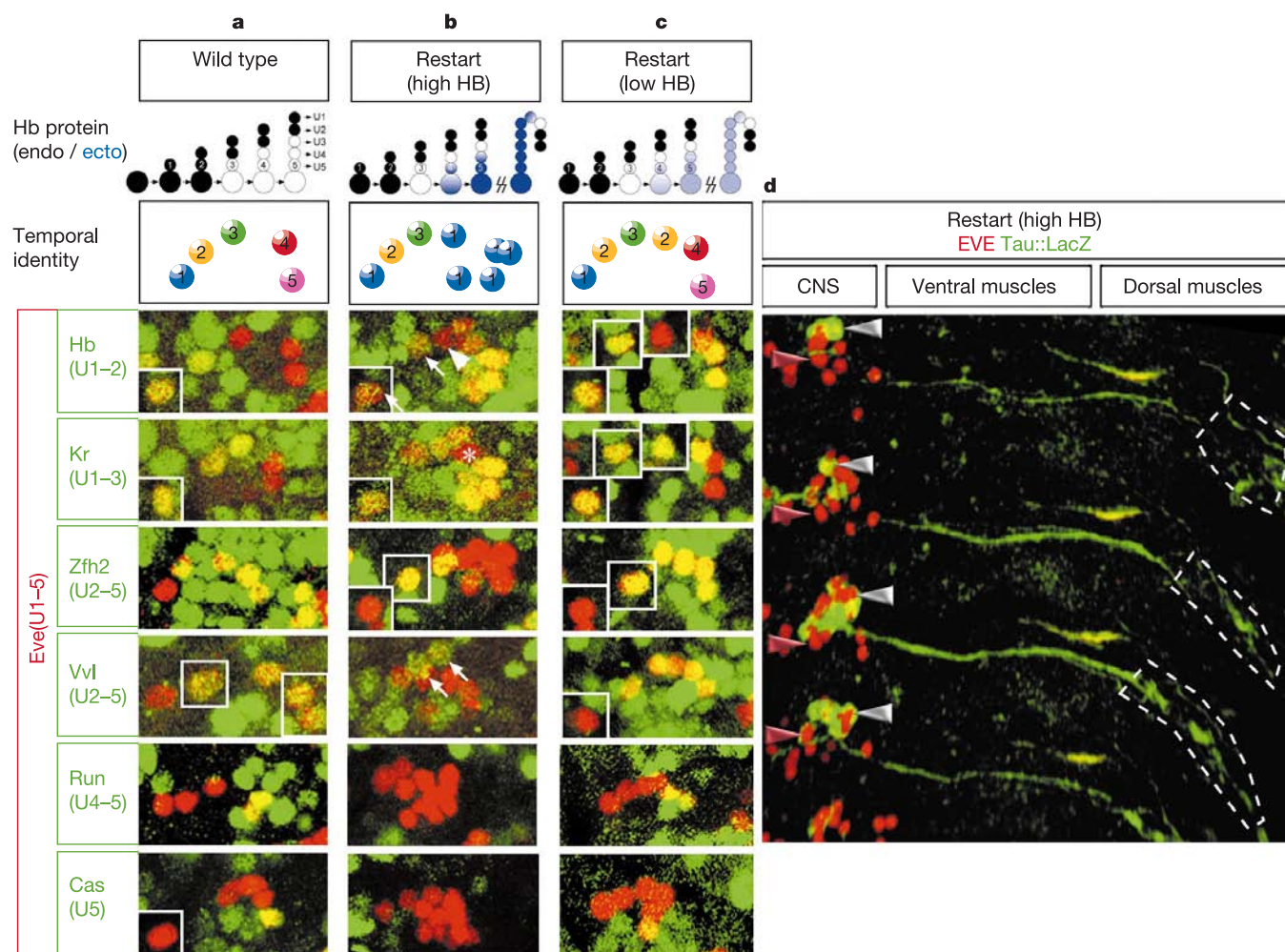


Figure 2 NB7-1 shows plasticity in response to Hunchback. In this figure and in Fig. 3, the top panels summarize Hb expression (endogenous Hb, black; ectopic Hb, blue), middle panels summarize U1–U5 temporal identity based on the data shown below, and bottom panels show the U1–U5 *Eve*⁺ neurons (red) stained for the indicated temporal identity markers (green). All panels show one hemi-segment of a stage-16 CNS (anterior, up; medial, left). **a**, Wild type. **b**, Re-expression of high levels of Hb in an older NB generates ectopic U1 neurons; the genotype is *+/+; prospero-Gal4/UAS-hb* at 23 °C. Ectopic Hb is not detected until after the birth of the Hb-negative U3 neuron (arrowhead), when it is observed in all later-born, lateral positioned U neurons (right side of panel). Asterisk indicates occasional Krüppel-negative U neuron of unknown identity. **c**, Re-expression of

low levels of Hb in an older NB generates ectopic presumptive U2 neurons; the genotype is *+/+; prospero-Gal4/UAS-hb* at 18 °C. Hb staining was imaged at high gain to clearly document the Hb-positive neurons; the Hb levels are much lower than in **b**. **d**, Ectopic U1 motor neurons project to dorsal muscles, similar to endogenous U1 motor neurons. The U1–U5 marker *Eve* is red, whereas the β gal from an *evef+3.5–4.3[tau::lacZ]* transgene showing mosaic expression in U motor neurons is green. The endogenous U1/U2 neurons do not express β gal (red arrowheads), but the β gal-positive ectopic U1 motor neurons (silver arrowheads) project to dorsal muscles. The transgene also shows ectopic body wall staining (dotted lines). Five segments are shown.

Table 1 Quantification of *hb* misexpression phenotypes in the NB7-1 lineage

Expt	Genotype*	Ectopic Hb levels†	Ectopic Hb timing‡	Number <i>Eve</i> ⁺ U neurons§	U neuron temporal identity					Conclusion
					U1	U2	U3	U4	U5	
1	Wild type	None	NA	5 (100)	1	1	1	1	1	Wild type
2	<i>pros-gal4</i> , 2 × <i>UAS-hb</i> (23 °C)	High	U4 →	9.3 (60)	7.3	1	1	0	0	Old NB is competent; high Hb = U1 fate
3	<i>pros-gal4</i> , 1 × <i>USA-hb</i> (18 °C)	Low	U4 →	6.1 (20)	1	2.1	1	1	1	Old NB is competent; low Hb = U2 fate
4	<i>en-gal4</i> , 2 × <i>USA-hb</i>	High	NB/GMC	16.4 (20)	16.4	0	0	0	0	Mitotic precursors competent
5	<i>eve-gal4</i> , 2 × <i>USA-hb</i>	High	Neurons	5 (30)	1	1	1	1	1	Post-mitotic neurons not competent
6	Wild type with heat shock	None	NA	5 (38)	1	1	1	1	1	Control for heat shock
7	<i>hs-hb</i> (3.5–5.5 h)	Low	U1/2	7.6 (30)	2.3	2.3	1	1	1	NB competent to make early-born fates
8	<i>hs-hb</i> (6.0–8.0 h)	Low	U3/4	5.9 (29)	1	1.9	1	1	1	NB less competent to make early-born fates
9	<i>hs-hb</i> (8.5–10 h)	Low	U4/5	5.4 (22)	1	1.4	1	1	1	NB less competent to make early-born fates
10	<i>hs-hb</i> (10.5–24 h)	Low	INs	5 (27)	1	1	1	1	1	NB not competent; temporal identity fixed

NA, not applicable.

*Genotype and temperature of experiment where relevant.

†Levels of ectopic Hunchback in lineage (qualitative).

‡Time of ectopic Hunchback in lineage (that is, U4 → indicates Hb expressed in NB7-1 as it generates GMC4/U4 and all subsequent progeny). INs, interneurons.

§Number per hemi-segment. Numbers in parentheses indicate number of hemi-segments scored.

||Average number of each cell type, including endogenous U1/U2 neurons, based on markers described in Fig. 2a.

The previous experiments induced Hb at a single point in the NB7-1 lineage, just after U3 is generated. To test the competence of NB7-1 at later points in its lineage, we use a heat-inducible *hsp70-hb* transgene¹⁸ to provide pulses of Hb at progressively later points in the lineage. When the Hb pulse is initiated early in the lineage, when U1 and U2 are normally generated, we detect several ectopic U1 and U2 neurons (Table 1, row 7). When Hb is provided later in the lineage, when U3 and U4 are normally generated, we observe just one ectopic presumptive U2 neuron (Table 1, row 8), and as Hb is

given progressively later, we observe a gradual decline in the production of presumptive U2 neurons, until no response to Hb is detected (Table 1, rows 9 and 10). Interestingly, an early pulse of Hb generates first-born U1 fates, but later pulses generate only presumptive U2 fates; perhaps only the early pulse, which combines induced and endogenous Hb levels, can achieve Hb levels sufficient to generate the U1 fate. We conclude that the NB shows a progressive restriction in its ability to generate early-born fates in response to Hb, consistent with models of progressive restriction proposed for vertebrate neural progenitors⁸.

We next address the question of when temporal identity is fixed during the process of neuronal differentiation. We begin by assaying NB and GMC competence, using *engrailed-Gal4* to limit Hb to NBs, GMCs and young neurons (Fig. 3a, 'early' panel), but not in mature post-mitotic neurons (Fig. 3a, 'late' panel), in the NB7-1 lineage. We observe a massive induction of U1 motor neurons, which stably maintain their U1 molecular marker profile through to the end of embryogenesis without detectable Hb protein (Fig. 3a and Table 1, row 4). To perform the converse experiment, we use *eve-Gal4* to express Hb specifically in post-mitotic U1–U5 motor neurons (Fig. 3b, cartoon). Despite high levels of Hb in these neurons, all U1–U5 cell fate markers remain wild type (Fig. 3b and Table 1, row 5). Similarly, *hsp70-hb* expression in newly post-mitotic neurons has no effect on their temporal identity (Table 1, rows 9 and 10). Thus, transient Hb expression in mitotic progenitors can induce early-born U1 identity, but permanent Hb in post-mitotic neurons does not alter their temporal identity. We currently cannot distinguish whether competence is lost in GMCs or in newly post-mitotic neurons. We conclude that mitotic progenitors, but not mature post-mitotic neurons, are competent to establish early-born neuronal identity in response to Hb.

Our results lead to two major conclusions. First, NBs exhibit both plasticity and progressive restriction during their cell lineage. Older NBs are competent to make early-born neurons in response to Hb, but this competence progressively declines over time (Fig. 4a). Some NBs may lose competence even more rapidly than NB7-1, however, as NB1-1 and NB4-2 show a more limited response to ectopic Hb¹¹. Second, only mitotic progenitors are competent to respond to Hb, and transient Hb in these progenitors is sufficient to establish heritable early-born neuron identity (Fig. 4b). This latter result may reflect the normal function of Hb in the CNS, because Hb is absent from *Eve*⁺ neurons by larval stages, and thus it may normally act to maintain *Eve* expression by inducing a heritable gene expression cascade or by an epigenetic mechanism. Interestingly, transient Hb leads to epigenetic silencing of *HOX* gene expression during *Drosophila* segmentation¹⁹, and the Hb-related Ikaros protein is required to silence gene expression in mature B cells during mammalian haematopoiesis²⁰. In both systems, the Hb/Ikaros proteins physically interact with Mi-2-Polycomb complex proteins

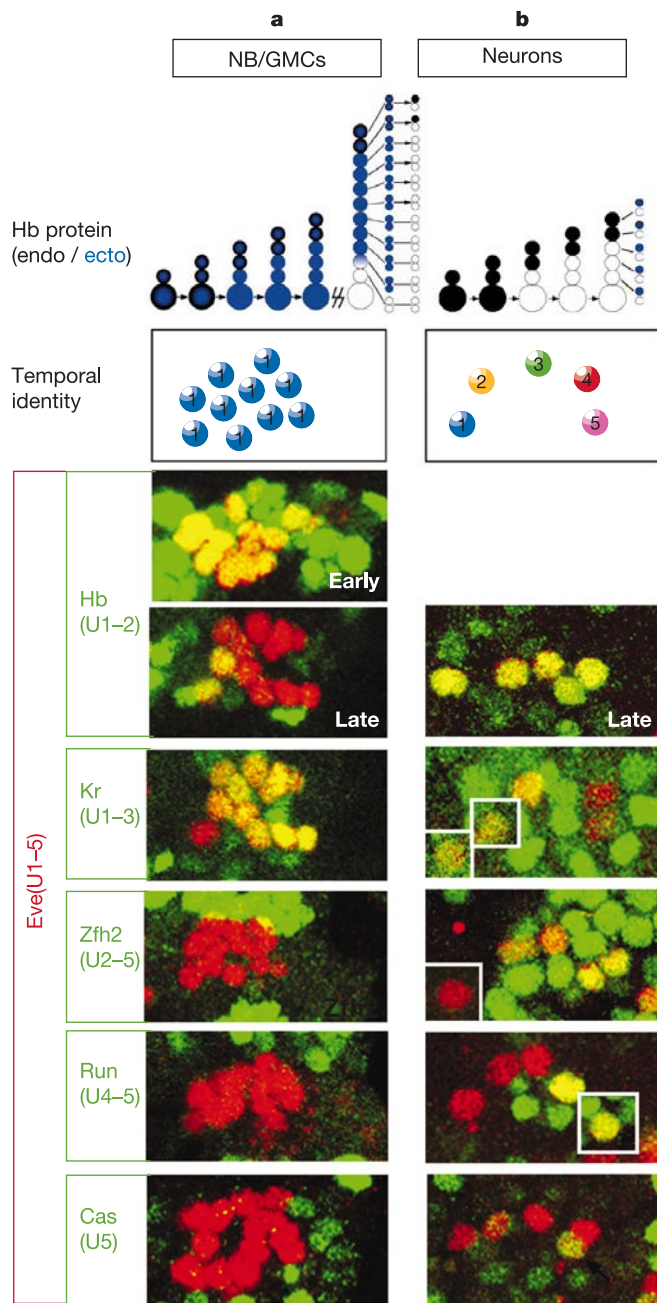


Figure 3 Mitotic progenitors but not post-mitotic neurons show competence to respond to Hunchback. **a**, Hb expression in NBs, GMCs and young neurons (early panel, stage 13 shown); the genotype is *engrailed-Gal4/UAS-hb*; +/*UAS-hb*. A large pool of mature U1 neurons maintain their temporal identity despite the absence of Hb protein (late panel, stage 17 shown). **b**, Hb expression in post-mitotic neurons; the genotype is *eve[+3.5-4.3]-Gal4/UAS-hb*; +/*UAS-hb*. Despite high Hb levels, all U1–U5 markers are expressed normally.

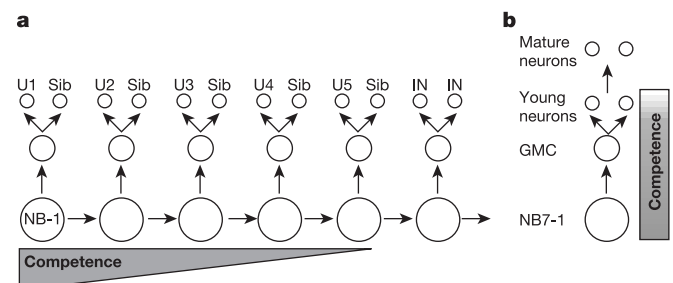


Figure 4 Summary of NB, GMC and neuronal competence to respond to Hb. **a**, NB7-1 shows progressive loss of competence to generate early-born neurons in response to Hb. **b**, The competence to generate early-born neurons in response to Hb is lost by the time a neuron becomes post mitotic.

to induce histone deacetylase-mediated gene silencing^{21,22}. This suggests an attractive model for temporal specification in the *Drosophila* CNS, in which Hb recruits Mi-2-Polycomb to silence genes conferring later-born temporal identity. The molecular basis of competence remains unclear, with one exception: we know that downregulation of Hb triggers progressive loss of competence, because maintaining continuous high Hb levels can indefinitely maintain competence¹¹. Is competence due to the presence of an unknown Hb target and cofactor, or the absence of a negative factor that suppresses early-born fates (such as Klumpfuss²³)? Does loss of competence in old NBs and in differentiating neurons occur by different mechanisms? Is there a distinct, later competence state for specifying subsequent Krüppel-positive temporal identities? *Drosophila* NBs now provide a model system for investigating the molecular nature of neural progenitor competence. □

Methods

We used the following fly stocks: (1) *hsp70-hb* (HB476.1 homozygous on chromosome III); (2) *yw; UAS-hb; UAS-hb*; (3) *yw; +; UAS-hb*; (4) *prospero-Gal4/prospero-Gal4* on chromosome III; (5) *prospero-Gal4/CyO, ftz-lacZ; eve-tau::lacZ/TM3, ftz-lacZ*; (6) *eve[+3.5-4.3]-Gal4/eve[+3.5-4.3]-Gal4* on chromosome II; (7) *engrailed-Gal4/engrailed-Gal4* on chromosome II; (8) *yw, hsp70-FLP* on chromosome X; and (9) *+/Act5c-FRT-stop-FRT-Tau::lacZ, CyO*.

For *hs-hb* experiments, embryos were collected for 1–2 h, aged to the indicated time, and subjected to three cycles of 30 min at 37 °C then 1 h at 22 °C, and allowed to develop to stage 16–17. This generates the maximum Hb level without affecting control embryos; Hb protein is detected throughout the CNS for 4 h after induction. Before heat shock, some embryos were fixed and stained with Eve/Hb for developmental staging. For *hs-FLP* experiments, embryos were subjected to heat shock for 22 min at 37 °C and aged to stage 16–17. Detailed methods available upon request.

All antibodies, staining procedures and imaging methods have been described previously¹¹, except for guinea-pig anti-Runt (1:500) and rat anti-Vvl (1:100) antibodies. All images were collected as confocal image stacks, processed in ImageJ (NIH), and shown as two-dimensional projections. U neurons are shown as insets in their approximate spatial position if they would be obscured in the projection.

Received 14 May; accepted 9 July 2003; doi:10.1038/nature01910.

- Turner, D. L. & Cepko, C. L. A common progenitor for neurons and glia persists in rat retina late in development. *Nature* **328**, 131–136 (1987).
- Perron, M., Kanekar, S., Vetter, M. L. & Harris, W. A. The genetic sequence of retinal development in the ciliary margin of the *Xenopus* eye. *Dev. Biol.* **199**, 185–200 (1998).
- Cepko, C. L., Austin, C. P., Yang, X., Alexiades, M. & Ezzeddine, D. Cell fate determination in the vertebrate retina. *Proc. Natl Acad. Sci. USA* **93**, 589–595 (1996).
- McConnell, S. K. The determination of neuronal fate in the cerebral cortex. *Trends Neurosci.* **12**, 342–349 (1989).
- Angevine, J. B. Jr & Sidman, R. L. Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. *Nature* **192**, 766–768 (1961).
- Qian, X. *et al.* Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. *Neuron* **28**, 69–80 (2000).
- Zhou, Q. & Anderson, D. J. The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. *Cell* **109**, 61–73 (2002).
- Desai, A. R. & McConnell, S. K. Progressive restriction in fate potential by neural progenitors during cerebral cortical development. *Development* **127**, 2863–2872 (2000).
- Livesey, F. J. & Cepko, C. L. Vertebrate neural cell-fate determination: lessons from the retina. *Nature Rev. Neurosci.* **2**, 109–118 (2001).
- Rapaport, D. H., Patheal, S. L. & Harris, W. A. Cellular competence plays a role in photoreceptor differentiation in the developing *Xenopus* retina. *J. Neurobiol.* **49**, 129–141 (2001).
- Isshiki, T., Pearson, B., Holbrook, S. & Doe, C. Q. *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511–521 (2001).
- Kambadur, R. *et al.* Regulation of POU genes by castor and hunchback establishes layered compartments in the *Drosophila* CNS. *Genes Dev.* **12**, 246–260 (1998).
- Bossing, T., Udolph, G., Doe, C. Q. & Technau, G. M. The embryonic central nervous system lineages of *Drosophila melanogaster*. I. Neuroblast lineages derived from the ventral half of the neuroectoderm. *Dev. Biol.* **179**, 41–64 (1996).
- Schmid, A., Chiba, A. & Doe, C. Q. Clonal analysis of *Drosophila* embryonic neuroblasts: neural cell types, axon projections and muscle targets. *Development* **126**, 4653–4689 (1999).
- Landgraf, M., Bossing, T., Technau, G. M. & Bate, M. The origin, location, and projections of the embryonic abdominal motoneurons of *Drosophila*. *J. Neurosci.* **17**, 9642–9655 (1997).
- Skeath, J. B. & Doe, C. Q. Sanpodo and Notch act in opposition to Numb to distinguish sibling neuron fates in the *Drosophila* CNS. *Development* **125**, 1857–1865 (1998).
- Buenzow, D. E. & Holmgren, R. Expression of the *Drosophila* gooseberry locus defines a subset of neuroblast lineages in the central nervous system. *Dev. Biol.* **170**, 338–349 (1995).
- Struhl, G., Johnston, P. & Lawrence, P. A. Control of the *Drosophila* body pattern by the hunchback morphogen gradient. *Cell* **69**, 237–249 (1992).
- Zhang, C. C. & Bienz, M. Segmental determination in *Drosophila* conferred by hunchback (hb), a repressor of the homeotic gene Ultrabithorax (Ubx). *Proc. Natl Acad. Sci. USA* **89**, 7511–7515 (1992).
- Sabbatini, P. *et al.* Binding of Ikaros to the lambda5 promoter silences transcription through a mechanism that does not require heterochromatin formation. *EMBO J.* **20**, 2812–2822 (2001).

- Kehle, J. *et al.* dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* **282**, 1897–1900 (1998).
- Kim, J. *et al.* Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity* **10**, 345–355 (1999).
- Yang, X., Bahri, S., Klein, T. & Chia, W. Klumpfuss, a putative *Drosophila* zinc finger transcription factor, acts to differentiate between the identities of two secondary precursor cells within one neuroblast lineage. *Genes Dev.* **11**, 1396–1408 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Posakony, M. Fujioka and J. Jaynes for fly stocks; W. Odenwald and M. Lundell for antibodies; and R. Grosskortenhaus, M. Freeman, M. Rolls, T. Isshiki, M. Westerfield, T. Brody and W. Odenwald for comments. This work was funded by NIH and HHMI, where C.Q.D. is an Investigator.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.Q.D. (cdoe@uoneuro.uoregon.edu).

Basal body dysfunction is a likely cause of pleiotropic Bardet–Biedl syndrome

Stephen J. Ansley^{1*}, Jose L. Badano^{1*}, Oliver E. Blacque^{3*}, Josephine Hill⁴, Bethan E. Hoskins^{1,4}, Carmen C. Leitch¹, Jun Chul Kim³, Alison J. Ross⁴, Erica R. Eichers⁵, Tanya M. Teslovich¹, Allan K. Mah³, Robert C. Johnsen³, John C. Cavender⁷, Richard Alan Lewis^{5,6}, Michel R. Leroux³, Philip L. Beales⁴ & Nicholas Katsanis^{1,2}

¹Institute of Genetic Medicine, and ²Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland 21287, USA

³Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6 Canada

⁴Molecular Medicine Unit, Institute of Child Health, University College London, London WC1N 1EH, UK

⁵Departments of Molecular and Human Genetics, and ⁶Departments of Ophthalmology, Pediatrics, and Medicine, Baylor College of Medicine, Houston, Texas 77030, USA

⁷King Khaled Eye Hospital, Riyadh, 11462, Saudi Arabia

* These authors contributed equally to this work

Bardet–Biedl syndrome (BBS) is a genetically heterogeneous disorder characterized primarily by retinal dystrophy, obesity, polydactyly, renal malformations and learning disabilities. Although five BBS genes have been cloned^{1–6}, the molecular basis of this syndrome remains elusive. Here we show that BBS is probably caused by a defect at the basal body of ciliated cells. We have cloned a new BBS gene, *BBS8*, which encodes a protein with a prokaryotic domain, *pilF*, involved in pilus formation and twitching motility. In one family, a homozygous null *BBS8* mutation leads to BBS with randomization of left–right body axis symmetry, a known defect of the nodal cilium. We have also found that *BBS8* localizes specifically to ciliated structures, such as the connecting cilium of the retina and columnar epithelial cells in the lung. In cells, *BBS8* localizes to centrosomes and basal bodies and interacts with PCM1, a protein probably involved in ciliogenesis. Finally, we demonstrate that all available *Caenorhabditis elegans* BBS homologues are expressed exclusively in ciliated neurons, and contain regulatory elements for RFX, a transcription factor that modulates the expression of genes associated with ciliogenesis and intraflagellar transport.

BBS exhibits substantial genetic heterogeneity and, although typically inherited in an autosomal recessive pattern, in some

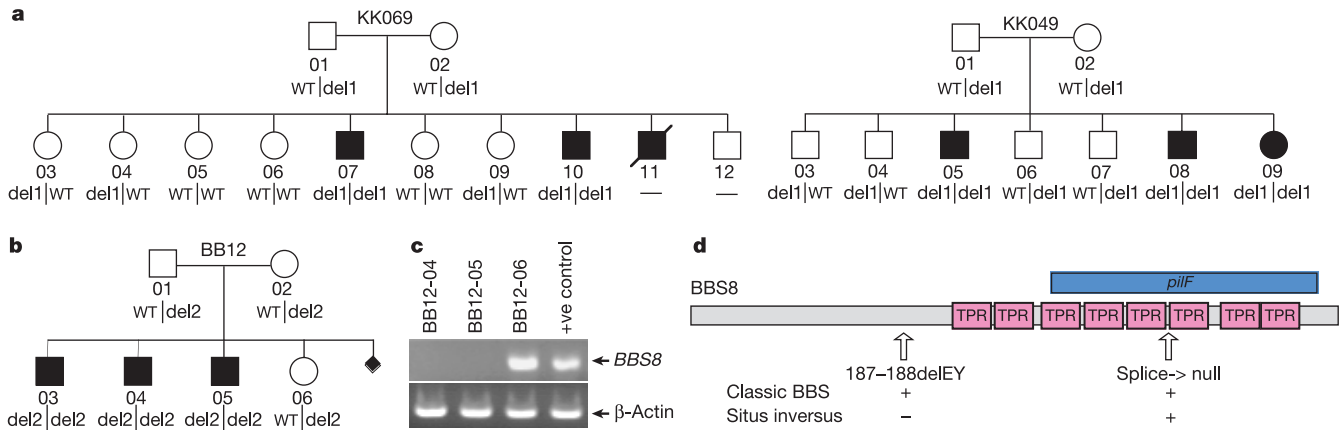


Figure 1 *BBS8* causes BBS. **a**, In families KK049 and KK069, affected individuals are homozygous for a 6-bp deletion that eliminates residues 187–188. Squares, males; circles, females. **b**, In pedigree BB12 all patients are homozygous for a 3-bp deletion that abolishes the splice donor site in intron 10. **c**, RT-PCR from tubular kidney cells cultured

from patients BB12-04, -05 and -06. *BBS8* messenger RNA is detected in the unaffected sister and the unrelated control but not the patients. **d**, Predicted domain structure of *BBS8* with TPRs depicted as boxes. The C terminus of *BBS8* is predicted to contain a *piIF* domain (blue box). The positions of the mutations are shown (arrows).

cases three mutant alleles in two genes are required for pathogenesis^{7–10}. Recently, a phylogenetic/genomic approach that identifies distant paralogues of known BBS proteins was successful in cloning new BBS loci¹. For the present study, we fragmented *BBS4* into eight overlapping segments and searched the conceptual translation of the draft human genome and dbEST. Because *BBS4* contains at least ten tetratricopeptide repeats (TPRs), we identified numerous TPR-containing transcripts. In one instance, however, we observed an alignment between three consecutive TPRs of *BBS4* and a contiguous region of the hypothetical protein *TTC8* (Supplementary Fig. 1a).

TTC8 maps to chromosome 14q32.11 (Supplementary Fig. 1b), a region that does not contain a mapped BBS locus. Nevertheless, on the basis of its similarity to *BBS4*, we hypothesized that *TTC8* might be involved in BBS. We tested this by screening *TTC8* in a cohort of 120 unrelated BBS patients. We identified homozygous alterations in patients from three families: further sequencing and clinical evaluations showed that all eight individuals affected in these families have a homozygous mutant genotype and exhibit classical BBS signs (Fig. 1a, b; see also Supplementary Fig. 2a and Table 1). In families KK049 and KK069 of Saudi Arabian lineage, each affected individual, but no unaffected relative, was homozygous for a 6-base-pair (bp) in-frame deletion in exon 6 that eliminates two amino acids (187–188delEY) (Fig. 1a; see also Supplementary Fig. 2a). We also detected a statistically significant association between the phenotype in these two families and *TTC8* at $\theta = 0$; two-point linkage analyses gave a LOD score from 3.36 (90% penetrance) to 3.54 (99% penetrance). Although the 187–188delEY allele might be in linkage disequilibrium with the true pathogenic mutation, we found either an E residue or a conservative substitution at position 187 in all 16 identifiable *TTC8* homologues—we observed similar conservation for the Y residue (14 of 16; 87.5%). More importantly, in all 16 species with the *TTC8* homologue, at least one of the two amino acids is conserved (Supplementary Fig. 2b). In a third consanguineous family of Pakistani descent (BB12) we identified a homozygous 3-bp deletion that abolishes the donor sequence at the splice junction of exon 10 (IVS10 + 2–4delTGC) in all three patients but not their unaffected sister (Fig. 1b). The expected effect of this mutation is a read-through culminating in an intronic stop codon. To substantiate this prediction, we cultured renal tubular cells and

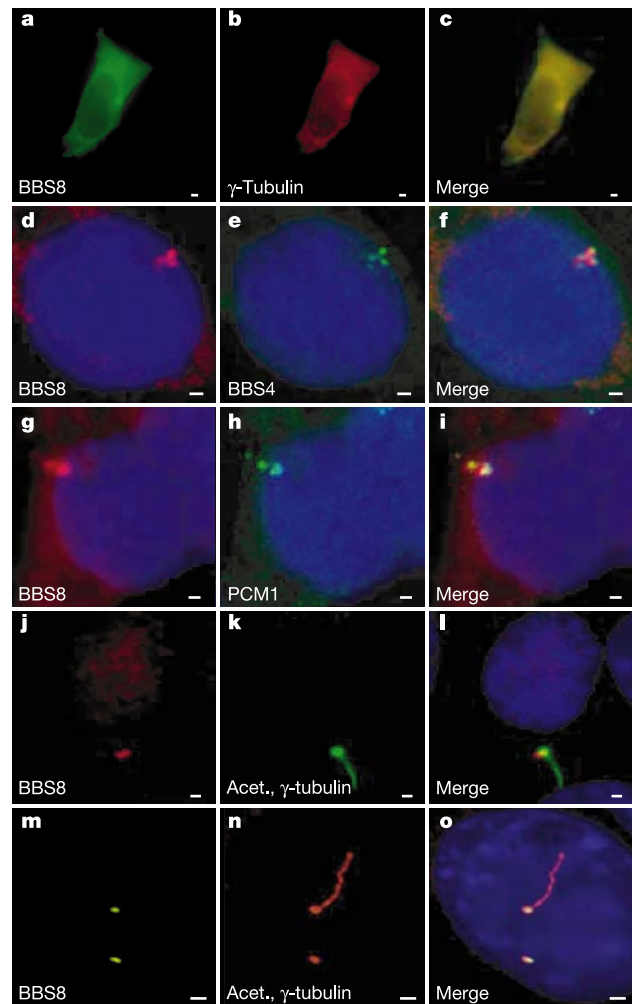


Figure 2 *BBS8* is a centrosomal and basal body protein. **a–i**, *BBS8* co-localizes with γ -tubulin (**a–c**; HEK293), *BBS4* (**d–f**; HeLa) and PCM1 (**g–i**; HeLa). **j–l**, In ciliated cells *BBS8* localizes to the basal body in NIH3T3 cells (**j–l**) and to both basal bodies and centrosomes in IMCD3 cells (**m–o**). Scale bar, 1 μ m. Acet., acetylated.

skin fibroblasts from patients BB12-04 and -05 and their unaffected sister BB12-06—who is heterozygous for the mutation—and performed polymerase chain reaction with reverse transcription (RT-PCR). We detected *BBS8* expression in BB12-06 and an unrelated control cell line, but not in the two BBS patients (Fig. 1c), suggesting that the aberrant *BBS8* transcript is probably eliminated by nonsense-mediated decay. These data, together with the absence of these alterations from 192 ethnically matched control chromosomes, led us to conclude that mutations in *TTC8* cause BBS, thus defining the eighth locus for this disorder, *BBS8*.

BBS8 generates two alternatively spliced isoforms (Supplementary Fig. 1b) and is widely expressed (Supplementary Fig. 3). The predicted protein contains eight TPRs towards the carboxy terminus and exhibits significant ($P < 10^{-5}$) similarity to a prokaryotic domain *pilF* (Fig. 1d; see also Supplementary Fig. 4), involved in twitching motility and type-IV pilus assembly^{11,12}, raising the possibility that *BBS8* might be relevant to the function of cilia,

flagella or pseudopodia. Notably, one of the patients in family BB12 manifests situs inversus, a known defect of left–right axis determination caused by dysfunction of the nodal cilium¹³. The association between BBS and the situs inversus phenotype is not coincidental (see Supplementary Fig. 5 and legend), although the presence of situs in only one of the three BB12 patients suggests that the defect is one of randomization of left–right symmetry.

To investigate *BBS8* further, we determined the localization of the protein in cells and tissues. Immunocytochemical analyses of HeLa and HEK293 cells identified *BBS8* as a centrosomal protein with a localization signal overlapping with γ -tubulin, an established centrosomal marker (Fig. 2a–c). This was reminiscent of the intracellular pattern of *BBS4* (Kim, J. C. *et al.*, manuscript submitted); double staining with antibodies against *BBS8* and *BBS4* showed co-localization near the centrosome (Fig. 2d–f). In ciliated NIH3T3 cells we found the protein at the base of the cilium (Fig. 2j–l), whereas in kidney IMCD3 cells, we found *BBS8* tightly

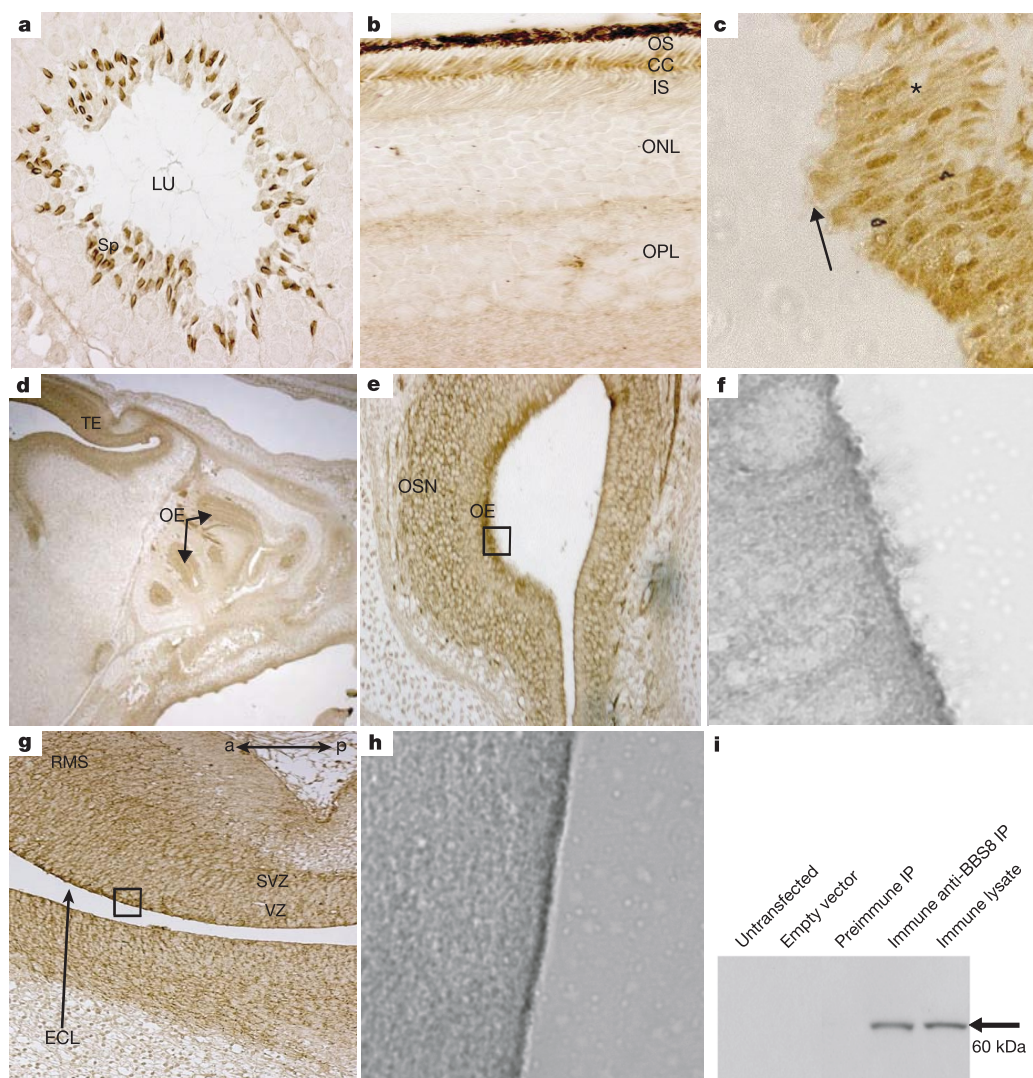


Figure 3 BBS8 localizes to ciliated structures in tissues. **a**, Maturing (stages X and XI), flagellated spermatids (Sp) surrounding the lumen (LU) of a 12 day (P12) mouse seminiferous tubule. **b**, Connecting cilium (CC) of a P12 mouse retina; no protein is detectable in outer or inner segments (OS, IS), outer nuclear layer (ONL), or outer plexiform layer (OPL). **c**, Adult human bronchial epithelium with prominent expression in ciliated columnar epithelial cells but not goblet cells (asterisk); the ciliary border is indicated (arrow). **d**, Sagittal section of an E14 mouse embryo showing BBS8 staining in

the olfactory epithelium (OE) and telencephalon (TE). **e**, Sagittal section of an E16 sinus, with BBS8 localizing to the olfactory epithelium (OE) and olfactory sensory neurons (OSN). **f**, Higher magnification of boxed area in **e**. **g**, BBS8 localizes to the ventricular and subventricular zones (VZ, SVZ) and the rostral migratory stream (RMS) of the anterior telencephalon (sagittal section, E16). Strong expression is visible in the ependymal cell layer (ECL, arrow). **h**, Higher magnification of box from **g**. **i**, Western blot showing specificity of the anti-BBS8 antibody. IP, immunoprecipitate.

associated with both the basal body and the centrosome, but not the ciliary axoneme (Fig. 2m–o).

We also queried whether BBS8 interacts with the other known BBS proteins or their interactors. Of particular interest was the BBS4-interacting protein PCM1 (Kim, J. C. *et al.*, manuscript submitted), a centrosomal protein also found in ciliary basal bodies^{14,15} and probably involved in centriolar replication during ciliogenesis¹⁵. Immunoprecipitations indicated that BBS8 binds to the C terminus of PCM1 (Supplementary Fig. 6). BBS8 also co-localizes with PCM1, confirming the immunoprecipitation data (Fig. 2g–i).

We next investigated the distribution of BBS8 in tissues. We raised a polyclonal antibody to human BBS8, which, based on its peptide sequence, is also expected to cross-react with its mouse orthologue. We detected specific staining of ciliated structures in 12-day-old mice (P12), including maturing (stages X and XI) spermatids

(Fig. 3a), the connecting cilium of the retina (Fig. 3b) and bronchial epithelial cells (Fig. 3c). In mouse embryos at 14 and 16 days old (E14 and E16), we also detected specific localization in the telencephalon, with prominent staining at the developing ependymal cell layer (which is ciliated) and the olfactory epithelium (Fig. 3d–h). Immunoprecipitation with the anti-BBS8 antiserum of protein extracts of Myc–BBS8-transfected HEK293 cells followed by western blotting revealed that the anti-BBS8 immune serum, but not the preimmune serum, immunoprecipitated specifically Myc–BBS8 (Fig. 3i).

To investigate further our hypothesis that BBS8 is associated with ciliary biogenesis or function, we used the nematode *C. elegans*, an important eukaryotic model system for ciliation and intraflagellar transport (IFT) studies¹⁶. Of the 959 somatic cells that make up an adult hermaphrodite, 302 are neuronal and only a subset (60) have ciliated dendritic endings.

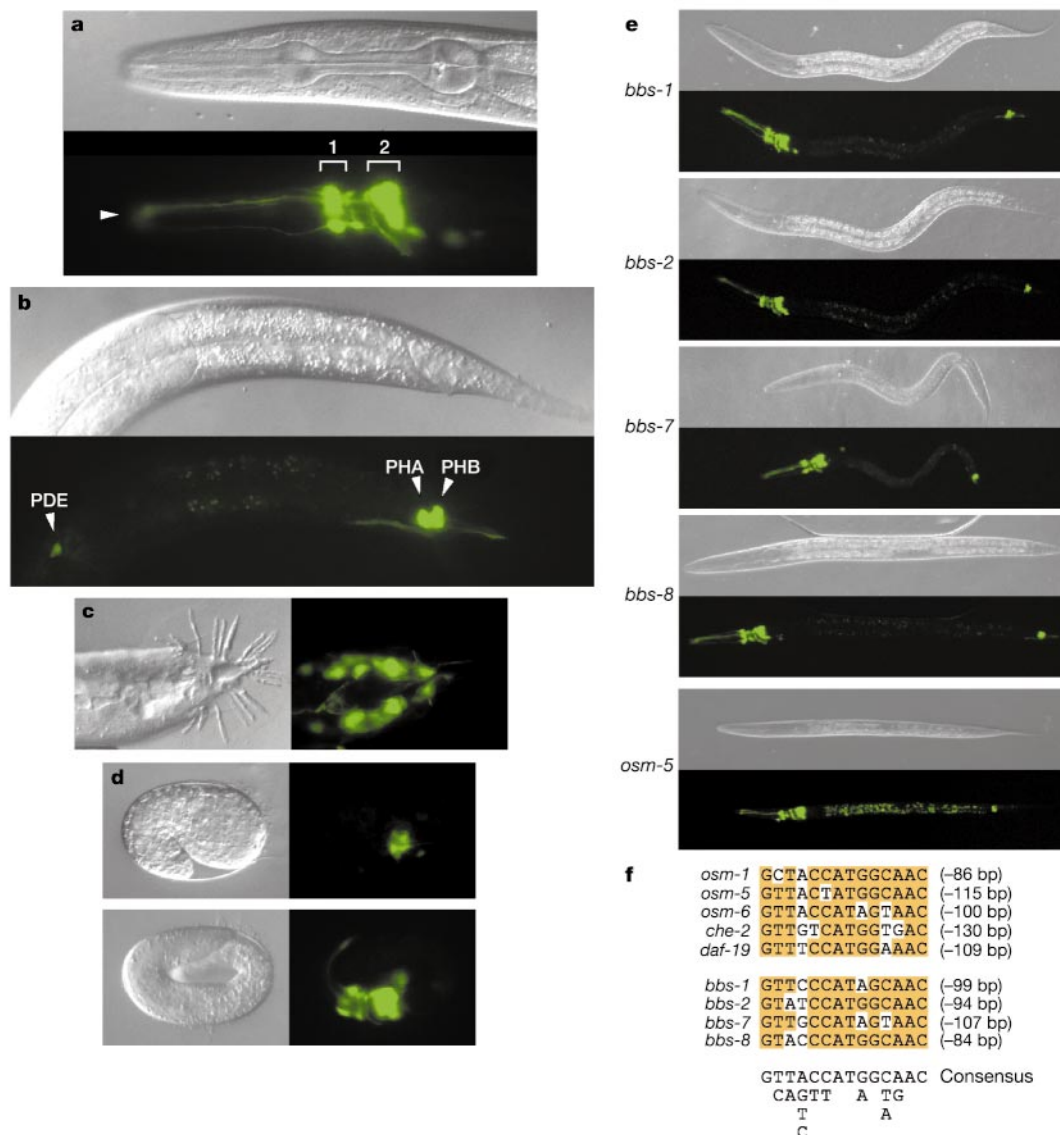


Figure 4 *Caenorhabditis elegans* *bbs* genes are expressed exclusively in ciliated neurons. **a–d**, DIC and GFP images of a nematode bearing a transcriptional *C. elegans* *bbs8::GFP* reporter (DIC image shown first). **a**, Head region of an L1/L2 animal showing strong expression in inner/outer labial neurons (1) and in amphid head neurons (2). Dendritic processes extend anteriorly to the nose (arrowhead). **b**, Tail region of an L3/L4 animal showing expression in PHA and PHB tail neurons and PDE neuronal cell. **c**, Adult male

showing expression in tail sensory ray neurons. **d**, *bbs-8* expression at the 1.5-fold (upper panel) and 3.0-fold (lower panel) embryonic stages. **e**, Transcriptional GFP reporter constructs of all *bbs* genes and *osm-5* in L1/L2 animals. **f**, Alignment of X boxes from *C. elegans* genes expressed in ciliated neurons and all *bbs* genes. Distances after the X boxes are from the translation start site (ATG) of each gene.

BLAST analysis of the *C. elegans* genome and proteome with BBS8 detected significant ($P < 10^{-8}$) similarity to a single predicted transcript, T25F10.5. To ascertain whether this gene is expressed in ciliated cells, we generated a transcriptional green fluorescent protein (GFP) expression construct and analysed transgenic lines carrying extrachromosomal arrays for GFP fluorescence. The expression pattern observed was indistinguishable from that of genes (for example, *osm-5*, the *C. elegans* orthologue of the mouse *Tg737* polycystic kidney disease gene¹⁷) whose spatial distribution is restricted to ciliated cells (Fig. 4a, b, e). Expression in hermaphrodites was most prominent in the head region (amphid and inner/outer labial neurons), in the tail region (in both phasmid neurons), and in the mid-body ciliated neuronal cell PDE (Fig. 4a, b). We also detected strong GFP signals in the male tail-ray neurons, all of which are ciliated (Fig. 4c). GFP expression was not detected in non-ciliated interneurons or motor neurons of the hermaphrodite tail or mid-body, or in any other tissues. The earliest point during development where expression was observed, at the 1.5-fold stage (Fig. 4d, upper panel), correlates with the onset of ciliogenesis¹⁸, and expression levels predictably increase as embryos develop to the threefold stage (Fig. 4d, lower panel).

Given that mutations in all known BBS genes give rise to identical or overlapping phenotypes, we postulated that all BBS proteins might be involved in the same cellular function. If this is true and the expression of the *C. elegans bbs-8* is relevant to the cellular mechanism of this disorder, other BBS orthologues should exhibit the same specific expression pattern. We investigated this by identifying the *C. elegans* orthologues of *BBS1*, *BBS2* and *BBS7* and determining their expression pattern. Consistent with our hypothesis, transgenic lines containing the three transcriptional *bbs::GFP* constructs showed expression patterns restricted exclusively to ciliated neurons and identical to that of *bbs-8* and *osm-5* (Fig. 4e).

Our data also suggest that each *C. elegans* BBS orthologue is regulated by the same mechanism. Computational analyses of the 5' untranslated regions (UTRs) of *bbs-1*, *bbs-2*, *bbs-7* and *bbs-8* identified a common regulatory element also present in numerous genes expressed in ciliated neurons (Fig. 4f). This motif, a 14-bp imperfect repeat termed X box, is centred typically about 100 bp upstream of the start codon¹⁹. The transcription factor regulating the expression of X boxes in *C. elegans*, DAF-19, is a member of the RFX protein family and is required for cilia formation¹⁹.

Collectively our data indicate a role for the BBS proteins in ciliary function. The BBS phenotype is consistent with this hypothesis, as many clinical aspects of this disorder can be explained by a ciliary defect. Dysfunction of the nodal cilium causes left-right axis defects in vertebrates^{13,20,21}. Also, compromised protein transport across the photoreceptor-connecting cilium causes retinal dystrophy^{21,22}, whereas failure of mechanosensation at the primary cilium of renal tubular cells causes polycystic kidney disease (PKD)²³. The best-studied mouse models of PKD, the Oak Ridge polycystic kidney mutant (*Tg737^{orp}*) and *cpk*, both develop BBS-like renal phenotypes, including renal cysts, with pancreatic and hepatic involvement^{21,24}.

Despite the phenotypic similarities with other ciliary defects, BBS is probably caused by a potentially new mechanism. BBS8 is not found in the axoneme and is therefore unlikely to be an IFT or structural axonemal protein. Rather, the specific localization of BBS8 to the basal body and its interaction with PCM1 raise the possibility that it either participates in ciliogenesis or mediates communication between the cilium and the interior of the cell. Intriguingly, BBS8 bears similarity to yeast *cdc23* (data not shown), a member of the anaphase-promoting complex. Inversin, another protein involved in ciliation and left-right axis determination, binds to the mammalian anaphase-promoting complex and has been postulated to be involved in the cell cycle^{25,26}. Our findings may thus facilitate the understanding of the link between ciliary function

and cellular response and provide an entry point to understanding the developmental and cognitive defects of BBS. □

Methods

Bardet-Biedl syndrome patients

R.A.L. and P.L.B. collated clinical data, performed clinical examinations and secured a diagnosis of BBS according to established criteria²⁷. Blood and clinical data were obtained with consent, in accordance with protocols approved by the human subjects ethics committees at each participating institution, and DNA was purified as described¹.

Mutation analyses

We determined the *BBS8* intron-exon structure by aligning its predicted complementary DNA sequence to the draft human genome as described¹. PCR products from BBS patients, relatives and ethnically matched controls were sequenced¹; primer sequences are available on request.

Expression studies

We performed RT-PCR and northern analyses with a 443-bp amplicon from the 3' UTR or a full-length cDNA clone of *BBS8* as described¹. Proximal tubular cells were selectively grown from urine samples in modified DMEM F12/10% FCS medium (Invitrogen) as were fibroblasts in HAM F10/12% FCS medium (GibcoBRL). After monolayer confluence, cells were trypsinized and RNA was prepared with Trizol (GibcoBRL) (1 ml cm⁻²). First-strand cDNA was synthesized by random priming with reverse transcriptase (Invitrogen). Complementary DNA was amplified with internal exonic primers corresponding to exons 9 and 13.

Immunocytochemistry

The open reading frame of human *BBS8* was cloned into the pCMV-Myc vector (BD Biosciences) and confirmed by double-strand sequencing. We cultured HeLa or HEK293 cells on 18-mm coverslips to 70–80% confluency and transfected them with a Myc-BBS8 plasmid or a GFP-BBS8 plasmid using the calcium phosphate kit (Invitrogen) or Polyfect transfection reagent (Qiagen). Cells were collected 24 h after transfection, fixed in -20 °C methanol, permeabilized in 0.1% Triton X-100, and blocked with phosphate buffer saline (PBS) containing 5.5% fetal bovine serum. We incubated the cells with anti-Myc mouse monoclonal antibody (Clontech). After washing with PBS, we incubated cells with a secondary rabbit anti-mouse antibody conjugated with Alexa Fluor 488 or 594 fluorophore (Molecular Probes). We used 4,6-diamidino-2-phenylindole to stain DNA, and coverslips were mounted in prolong antifade reagent (Molecular Probes). For triple-staining experiments, we used a γ -tubulin monoclonal antibody (Molecular Probes) or a PCM1 rabbit polyclonal antibody (a gift from A. Merdes) and a goat anti-rabbit secondary antibody conjugated with Alexa Fluor 488 or 594 dyes. Cilia were visualized with an acetylated tubulin monoclonal antibody (Sigma).

Antibodies and immunohistochemistry

We generated a polyclonal BBS8 antibody by injecting rabbits with the amino terminal (position 111–126) peptide GFLRPSTQSGRPGTME. This antibody was used in immunohistochemistry of 5- μ m-thick sagittal sections from mouse embryos and adult mouse and human tissues (Novagen) according to standard protocols. The primary anti-BBS8 antibody was used in 1:1,000 and 1:2,000 dilutions in block, applied to the slides and incubated overnight at 4 °C. Staining was detected with the ABC complex (DAKO) and DAB (Biomed) according to the manufacturer's instructions.

Immunoprecipitations

HEK293 cells grown in 100-mm tissue culture dishes were transfected transiently with Myc-BBS8 and haemagglutinin-PCM1 (encoding the C terminus of the protein from amino acid 1574 to the end) with the calcium phosphate kit (Invitrogen). Cells were collected 48 h after transfection in co-immunoprecipitation buffer (150 mM NaCl; 50 mM Tris-HCl pH 7.5; 1% NP-40) supplemented with protease inhibitor (Roche) and 500 μ l of 100 mM sodium orthovanadate. Cell lysates were incubated overnight with 10 μ g anti-Myc monoclonal antibody (immunoprecipitate) immobilized onto Sepharose beads (Covance). The immunoprecipitates were washed three times with co-immunoprecipitation buffer, resuspended in loading buffer and used in western blots. Immunoblot PVDF membranes (Bio-Rad) were blocked with 0.2% Tween 20 in PBS and 5% milk, probed with an anti-haemagglutinin mouse monoclonal antibody coupled to peroxidase (Roche), and detection was performed using an ECL system (Amersham-Pharmacia).

Caenorhabditis elegans bbs::GFP expression constructs

Transcriptional GFP expression constructs, generated as PCR products using a PCR-fusion-based approach²⁸, were constructed by placing the upstream UTR of *C. elegans* genes Y105E8A.5 (*bbs-1*), F20D12.3 (*bbs-2*), Y75B8A.12 (*bbs-7*) and T25F10.5 (*bbs-8*) 5' to a DNA fragment containing GFP and the UNC-54 3' UTR. Specifically, the *bbs* constructs contain 500 bp of upstream sequence and the first 15 bp of exon 1 for *bbs-1*; 251 bp UTR and first 14 bp of exon 1 for *bbs-2*; 1,559 bp UTR and first 15 bp of exon 1 for *bbs-7*; and 370 bp UTR and first 16 bp of exon 1 for *bbs-8*. The GFP-UNC-54 3' UTR fragment, which also contains a nuclear localization signal sequence 5' to GFP, was amplified from the GFP expression vector, pPD95.67, provided by A. Fire. Transgenic lines carrying extrachromosomal arrays of the *bbs::GFP* expression constructs were generated by coinjection of *dpy-5(e907)* nematodes with the GFP construct and a rescue plasmid, pCeh 341, which contains wild-type *dpy-5* (provided by C. Thacker and A. Rose). The *osm-5::GFP* construct was obtained from B. K. Yoder. All nematodes were cultured as described previously²⁹.

Received 24 June; accepted 8 September 2003; doi:10.1038/nature02030.
Published online 21 September 2003.

1. Badano, J. L. *et al.* Identification of a novel Bardet-Biedl syndrome protein, BBS7, that shares structural features with BBS1 and BBS2. *Am. J. Hum. Genet.* **72**, 650–658 (2003).
2. Slavotinek, A. M. *et al.* Mutations in MKKS cause Bardet-Biedl syndrome. *Nature Genet.* **26**, 15–16 (2000).
3. Katsanis, N. *et al.* Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. *Nature Genet.* **26**, 67–70 (2000).
4. Nishimura, D. Y. *et al.* Positional cloning of a novel gene on chromosome 16q causing Bardet-Biedl syndrome (BBS2). *Hum. Mol. Genet.* **10**, 865–874 (2001).
5. Myktyyn, K. *et al.* Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nature Genet.* **31**, 435–438 (2002).
6. Myktyyn, K. *et al.* Identification of the gene that, when mutated, causes the human obesity syndrome BBS4. *Nature Genet.* **28**, 188–191 (2001).
7. Beales, P. L. *et al.* Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. *Am. J. Hum. Genet.* **72**, 1187–1199 (2003).
8. Katsanis, N. *et al.* Triallelic inheritance in Bardet-Biedl syndrome, a mendelian recessive disorder. *Science* **293**, 2256–2259 (2001).
9. Badano, J. L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nature Rev. Genet.* **3**, 779–789 (2002).
10. Katsanis, N. *et al.* BBS4 is a minor contributor to Bardet-Biedl syndrome and may also participate in triallelic inheritance. *Am. J. Hum. Genet.* **71**, 22–29 (2002).
11. Maurer, L. & Orndorff, P. E. Identification and characterization of genes determining receptor binding and pilus length of *Escherichia coli* type 1 pili. *J. Bacteriol.* **169**, 640–645 (1987).
12. Merz, A. J., So, M. & Sheetz, M. P. Pilus retraction powers bacterial twitching motility. *Nature* **407**, 98–102 (2000).
13. Nonaka, S. *et al.* Randomization of left-right symmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell* **95**, 829–837 (1998).
14. Dammermann, A. & Merdes, A. Assembly of centrosomal proteins and microtubule organization depends on PCM-1. *J. Cell Biol.* **159**, 255–266 (2002).
15. Kubo, A., Sasaki, H., Yuba-Kubo, A., Tsukita, S. & Shiina, N. Centriolar satellites: Molecular characterization, ATP-dependent movement toward centrioles and possible involvement in ciliogenesis. *J. Cell Biol.* **147**, 969–979 (1999).
16. Rosenbaum, J. L. & Witman, G. B. Intraflagellar transport. *Nature Mol. Cell Biol.* **3**, 813–825 (2002).
17. Haycraft, C. J., Swoboda, P., Taulman, P. D., Thomas, J. H. & Yoder, B. K. The *C. elegans* homolog of the murine cystic kidney disease gene *Tg737* functions in a ciliogenic pathway and is disrupted in *osm-5* mutant worms. *Development* **128**, 1493–1505 (2001).
18. Fujiwara, M., Ishihara, T. & Katsura, I. A novel WD40 protein, CHE-2, acts cell-autonomously in the formation of *C. elegans* sensory cilia. *Development* **126**, 4839–4848 (1999).
19. Swoboda, P., Adler, H. T. & Thomas, J. H. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. *Mol. Cell* **5**, 411–421 (2000).
20. Okada, Y. *et al.* Abnormal nodal flow precedes situs inversus in *iv* and *inv* mice. *Mol. Cell* **4**, 459–468 (1999).
21. Pazour, G. J. *et al.* The intraflagellar transport protein, IFT88, is essential for vertebrate photoreceptor assembly and maintenance. *J. Cell Biol.* **157**, 103–113 (2002).
22. Marszalek, J. R. *et al.* Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors. *Cell* **102**, 175–187 (2000).
23. Nauli, S. M. *et al.* Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. *Nature Genet.* **33**, 129–137 (2003).
24. Hou, X. *et al.* Cystin, a novel cilia-associated protein, is disrupted in the *cpk* mouse model of polycystic kidney disease. *J. Clin. Invest.* **109**, 533–540 (2002).
25. Morgan, D. *et al.* Expression analyses and interaction with the anaphase promoting complex protein Apc2 suggest a role for inversin in primary cilia and involvement in the cell cycle. *Hum. Mol. Genet.* **15**, 3345–3350 (2002).
26. Watanabe, D. *et al.* The left-right determinant Inversin is a component of node monocilia and other 9 + 0 cilia. *Development* **130**, 1725–1734 (2003).
27. Beales, P. L., Elcioglu, N., Woolf, A. S., Parker, D. & Flintner, F. A. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J. Med. Genet.* **36**, 437–446 (1999).
28. Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques* **32**, 728–730 (2002).
29. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank all the BBS families for their willing and continued participation in our studies; J. Sowden, S. Darling and R. Graham for technical help; and J. Lupski, A. Chakravarti, J. Nathans, P. Scambler, A. McCallion and L. Kotch for their critical evaluation of this manuscript. We also thank J. Morton for clinical details and J. Goodship for discussions. This study was supported by grants from the National Institute of Child Health and Development, NIH and the March of Dimes (N.K.), the Research Department of the King Khaled Eye Specialist Hospital, Riyadh, Saudi Arabia (J.C.C., R.A.L.), the Foundation Fighting Blindness, USA (R.A.L.), the Research to Prevent Blindness, New York (R.A.L.), NCIC, HSFC&Y, CIHR and MSFHR (M.R.L.), NSERC (R.C.J.), Genome BC and Genome Canada (R.C.J., K.M.), the National Kidney Research Fund (B.E.H.), the Birth Defects Foundation (P.L.B.), and the Wellcome Trust (P.L.B.).

Authors' contributions The laboratories of M.R.L., P.L.B. and N.K. contributed equally to this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.K. (katsanis@jhmi.edu). Nucleotide sequences for the two BBS8 splice isoforms (AY366523 (long isoform) and AY366524 (short isoform)) have been deposited in GenBank.

The Wnt/ β -catenin pathway regulates cardiac valve formation

Adam F. L. Hurlstone*, Anna-Pavlina G. Haramis*, Erno Wienholds, Harry Begthel, Jeroen Korving, Fredericus van Eeden, Edwin Cuppen, Danica Zivkovic, Ronald H. A. Plasterk & Hans Clevers

Netherlands Institute for Developmental Biology, Hubrecht Laboratory and Centre for Biomedical Genetics, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands

* These authors contributed equally to this work

Truncation of the tumour suppressor adenomatous polyposis coli (Apc) constitutively activates the Wnt/ β -catenin signalling pathway¹. Apc has a role in development: for example, embryos of mice with truncated Apc do not complete gastrulation². To understand this role more fully, we examined the effect of truncated Apc on zebrafish development. Here we show that, in contrast to mice, zebrafish do complete gastrulation. However, mutant hearts fail to loop and form excessive endocardial cushions. Conversely, overexpression of Apc or Dickkopf 1 (Dkk1), a secreted Wnt inhibitor³, blocks cushion formation. In wild-type hearts, nuclear β -catenin, the hallmark of activated canonical Wnt signalling⁴, accumulates only in valve-forming cells, where it can activate a Tcf reporter. In mutant hearts, all cells display nuclear β -catenin and Tcf reporter activity, while valve markers are markedly upregulated. Concomitantly, proliferation and epithelial-mesenchymal transition, normally restricted to endocardial cushions, occur throughout the endocardium. Our findings identify a novel role for Wnt/ β -catenin signalling in determining endocardial cell fate.

Apc is an essential component of the axin-containing destruction complex that phosphorylates β -catenin, tagging it for ubiquitination and degradation by the proteasome. In the presence of a Wnt ligand, β -catenin is stabilized and accumulates in the nucleus where it binds and activates Tcf transcription factors¹. APC mutations, common in colorectal cancer, occur proximal to the axin-binding motifs in the mutation cluster region (MCR; Fig. 1a). These truncations lead to constitutive activation of the pathway.

We have recently developed a reverse genetics strategy for inactivating genes in the zebrafish germline⁵. The current zebrafish genome database contains a single *apc* orthologue (Supplementary Fig. 1a, b). We screened an F₁ N-ethyl-N-nitrosourea (ENU)-mutagenized zebrafish library for *apc* nonsense mutations mapping to the putative MCR. A premature stop codon corresponding to amino acid (a.a.) 1318 of human APC was identified. The allele was designated *apc*^{mcr}, and is predicted to constitutively activate Wnt/ β -catenin signalling.

apc^{mcr} heterozygotes developed normally. Intercrossing resulted in clutches of F₃ embryos of which 25% died between 72 and 96 hours post-fertilization (h.p.f.), displaying multiple defects. These included, most prominently, cardiac malformation with associated pericardial oedema and blood pooling (Fig. 1b), enlarged otic vesicles, smaller eyes, and body curvature. Further, jaw, pharynx, and inner-ear structures failed to develop and fin buds arrested. Primordia for internal organs such as gut, liver and pancreas formed but developed abnormally (A.F.L.H. and A.P.G.H., unpublished observations). Genotyping revealed complete correspondence between this phenotype and homozygosity for the *apc*^{mcr} mutation. Mutant embryos probably developed beyond gastrulation owing to the presence of maternal Apc (data not shown).

To verify that the above developmental defects were due to loss of Apc function and not to co-segregation of an unidentified linked mutation, we injected zygotes resulting from intercrosses

of *apc*^{mcr/+} heterozygotes with 200 pg of RNA encoding a human APC fragment (a.a. 1020–2032) containing the β -catenin and axin binding domains fused to green fluorescent protein (GFP). We observed a 66% reduction in the expected number of mutants at 48 h.p.f. (23/271 compared to an anticipated 68/271), whereas the expected frequency of mutants was observed in non-injected siblings (24.6%; $n = 167$). Genotyping confirmed that 59.1% of *apc*^{mcr/mcr} homozygotes now appeared phenotypically normal, whereas another 12.5% had normal hearts, but retained most other defects (for example, enlarged otic vesicles, smaller eyes, and body curvature) (Fig. 2c). This implied that the mutant phenotype was due specifically to loss of the Wnt-regulatory function of Apc, and that the heart defects were not secondary to other abnormalities.

In an independent forward genetic screen (D.Z., to be published elsewhere), we identified a second ENU-induced mutant *apc* allele (termed *CA50a*), whose phenotype was indistinguishable from that of the *mcr* mutant (Fig. 1b, d). The *CA50a* allele failed to complement the *mcr* allele. The *CA50a* mutation results in premature stop codon truncation of the encoded gene product at a Leu residue corresponding to position 613 of the human protein.

The heart is the first organ to form and function during vertebrate embryogenesis, and cardiac malformation was the earliest gross developmental defect exhibited by *apc* mutants. Heart morphology, expression of the cardiac marker *nkx2.5*, and chamber specification (ascertained by expression of *cmlc2* and *vmhc*) were all normal at 36 h.p.f. in mutants (data not shown). Mutant hearts comprised both myocardial and endocardial cell layers and initially manifested vigorous, rhythmic peristaltic contractions. Subsequently, however, they failed to undergo looping morphogenesis and contractile function diminished progressively, such that by around 80 h.p.f. both chambers became silent and blood circulation

ceased altogether. Prior to this, blood was observed regurgitating within mutant heart chambers (Supplementary videos), indicative of a valve defect. Histological examination and Nomarski microscopy revealed that the discrete endocardial cushions (precursors of the valves proper) positioned between the atrium and ventricle in wild-type hearts had been replaced by a profuse endocardial layer fused at the atrioventricular (AV) boundary in mutant hearts. All endocardial cells appeared to have undergone epithelial–mesenchymal transition (Fig. 1c, d).

In the APC–GFP RNA injection experiment (see above), we observed a class of embryos (6.6%, 18/271) in which hearts failed to undergo looping and blood regurgitated, but that were otherwise morphologically normal at 48 h.p.f. Phase contrast microscopy and histology revealed the absence of endocardial cushions (not shown). These embryos were genotypically wild type (Fig. 2c). Injection of wild-type zygotes with 500 pg of APC–GFP RNA increased the fraction of embryos lacking endocardial cushions at 48 h.p.f. to 11.2% (25/224). This implied that blocking endogenous Wnt/ β -catenin signalling could inhibit endocardial cushion formation. To further validate this hypothesis, we injected wild-type embryos with 20 pg of Dkk1 RNA. As reported⁶, this resulted in forebrain expansion accompanied by a mild reduction in trunk and tail tissue. Cardiogenesis and vasculogenesis, however, were not compromised (our findings). At 48 h.p.f., we observed a lack of heart looping and blood regurgitation in 33.5% (106/316) of Dkk1-injected embryos, but not in controls (Fig. 2a). Endocardial cushions were absent (Fig. 2b).

Immunohistochemical staining for β -catenin revealed nuclear β -catenin, indicative of activated canonical Wnt signalling⁴ in endocardial cells populating the AV cushions and cushions of the bulbous arteriosus of wild-type hearts and the myocardial cells immediately overlying them (Fig. 3a and Supplementary Fig. 2). In

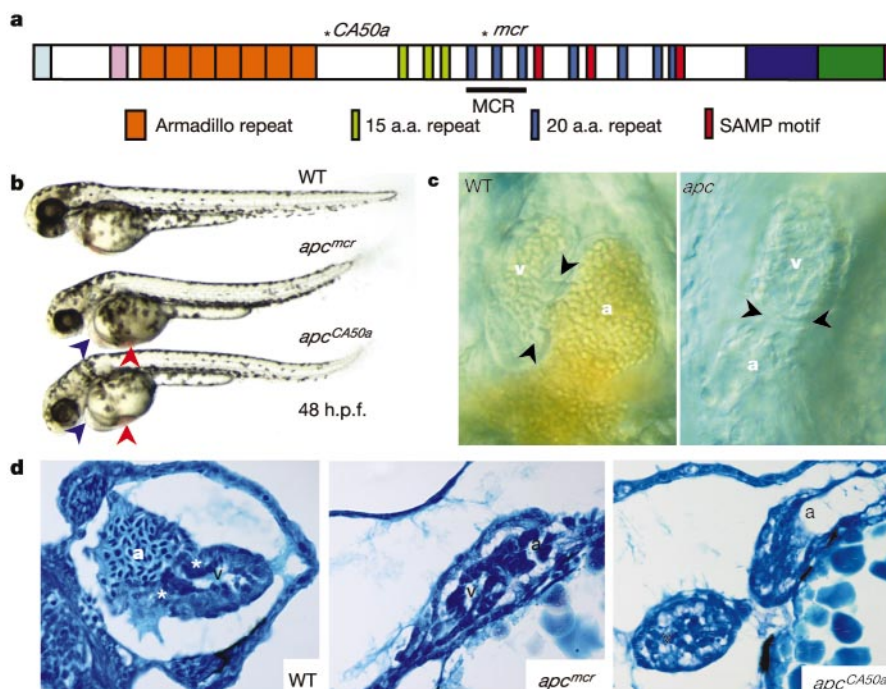


Figure 1 Mutation of *apc* results in heart malformation. **a**, Diagram of wild-type (WT) Apc protein; 15 and 20 a.a. repeats mediating β -catenin binding, and SAMP motifs mediating Axin binding, are depicted. Asterisks indicate the position of the *mcr* and *CA50a* truncations. a.a., amino acid; MCR, mutation cluster region. **b**, Morphology of WT, *apc*^{mcr/mcr} and *apc*^{CA50a/CA50a} embryos at 48 h.p.f. Blue arrowheads indicate pericardial oedema and red arrowheads pooled blood in mutants. **c**, Nomarski microscopy of WT and

apc hearts at 72 h.p.f. Arrowheads indicate the endocardial cushions in the WT and the fused endocardial layer at the AV boundary in *apc*^{mcr/mcr}. **d**, Transverse section of WT and sagittal sections of *apc*^{mcr/mcr} and *apc*^{CA50a/CA50a} hearts at 72 h.p.f. Asterisks indicate endocardial cushions. Note in *apc* hearts the profuse endocardium occluding the aperture between heart chambers and excess cardiac jelly. a, Atrium; v, ventricle.

contrast, all endocardial and myocardial cells within mutant hearts displayed prominent nuclear β -catenin (Fig. 3b). *apc^{mcr}* carriers were crossed with a transgenic zebrafish line expressing GFP under a Tcf responsive promoter (TOPdGFP)⁷. In wild-type embryos we observed GFP expression only within endocardial cushions (Fig. 3c). In *apc* mutant hearts, GFP expression occurred throughout the endo- and myocardium (Fig. 3d). Expression of proliferating cell nuclear antigen (PCNA) mirrored the pattern of nuclear β -catenin, being restricted to transdifferentiated endocardial cells in cushions (Fig. 3e). In mutant hearts, all endocardial cells were PCNA positive (Fig. 3f).

We next examined the hearts of *axin1* mutant (*mbl*) zebrafish^{8,9}. Pericardial oedema and blood pooling were observed in the majority (81%; 50/62) of *mbl* mutants. Frequently, this was accompanied by reduced (26%) or absent (13%) looping of the heart tube and blood regurgitation within heart chambers (complete loss of circulation was observed in 7% of *mbl* embryos by 72 h.p.f.). The most severely affected *mbl* mutant hearts closely resembled *apc* mutant hearts (data not shown). *mbl* hearts with intermediate looping displayed enlarged endocardial cushions and a concomitant increase in nuclear β -catenin and PCNA (Supplementary Fig. 3).

To further investigate the link between Wnt/ β -catenin signalling and valve formation, we performed *in situ* hybridization (ISH) for *bmp4*, *versican* (or *cspg2/br146*) and *notch1b*. As reported previously¹⁰ all three genes are initially expressed throughout the anteroposterior extent of the heart, *bmp4* and *versican* in the myocardium and *notch1b* in the endocardium. Later, expression of these genes is restricted to the AV valve-forming region—*bmp4* and *versican* by 37 h.p.f., and *notch1b* by 45 h.p.f.. At 36 h.p.f., the expression pattern of these three markers was indistinguishable between mutant and wild-type hearts (not shown). By 48 h.p.f., *bmp4* and *versican* expression was restricted to a few myocardial cells at the valve-forming region in wild-type hearts (Fig. 4a, c). In mutant hearts, both genes were dramatically upregulated and the domains of expression greatly expanded, encompassing the entire ventricle (Fig. 4b, d). Likewise, endocardial expression of *notch1b*

was valve-specific in wild-type hearts at 48 h.p.f. (Fig. 4e), but a broader expression domain was detected in mutant hearts (Fig. 4f). At 72 h.p.f., *notch1b* expression was observed throughout the endocardium of mutant hearts (inset Fig. 4f). Another endocardial valve-specific marker, *hyaluronan synthase 2* (*has2* or *DG42*) (J. Bakkers, personal communication) was restricted to the AV valve-forming region in both wild-type and mutant hearts at 48 h.p.f., albeit already upregulated in mutant hearts (insets Fig. 4g, h). At 72 h.p.f., *has2* expression remained restricted in wild-type hearts (Fig. 4g) but was expressed throughout the endocardium of mutants (Fig. 4h). As *bmp4*, *versican* and *has2* are purportedly transcriptional targets of the Wnt/ β -catenin pathway (refs 4, 11, 12, and M. Morkel and W. Birchmeier, personal communication), these data confirmed that Wnt/ β -catenin signalling is operative in myocardial and endocardial cells only at the valve-forming region in wild-type hearts.

Cardiac valve formation depends on signalling between myocardial and endocardial cell layers across an elaborate extracellular matrix, involving TGF- β , BMP, and EGF family members^{13–15}. Here we uncover a role for the Wnt/ β -catenin pathway in this network. Wnt/ β -catenin signalling is probably not involved in valve specification. Rather, it regulates subsequent expression of valve markers as well as proliferation and transdifferentiation of endocardial cells to establish endocardial cushions (model in Supplementary Fig. 4). Mice mutant in the Wnt target genes *versican* or *has2* fail to develop endocardial cushions^{16,17}. Similarly, *jekyll*

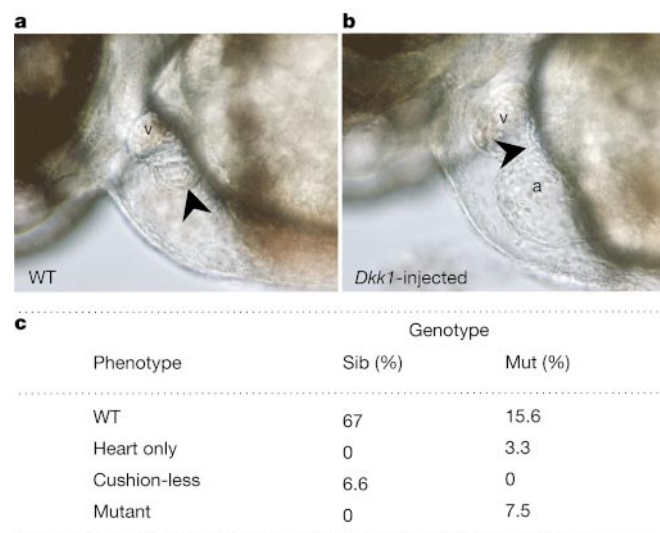


Figure 2 Wnt/ β -catenin signalling regulates endocardial cushion formation. Lateral views of hearts at 48 h.p.f. in a WT (a) and *Dkk1*-injected embryo (b). Arrowheads indicate one of the two AV endocardial cushions in a and endocardial monolayer at the AV boundary in b. a, atrium; v, ventricle; WT, wild type (non-injected control). Anterior to the left. Original magnification $\times 200$. c, Percentage of embryos with a given phenotype and genotype following injection of RNA encoding APC-GFP. 'Heart only' denotes rescue of only the heart defect, while 'cushion-less' indicates inhibition of endocardial cushion formation.

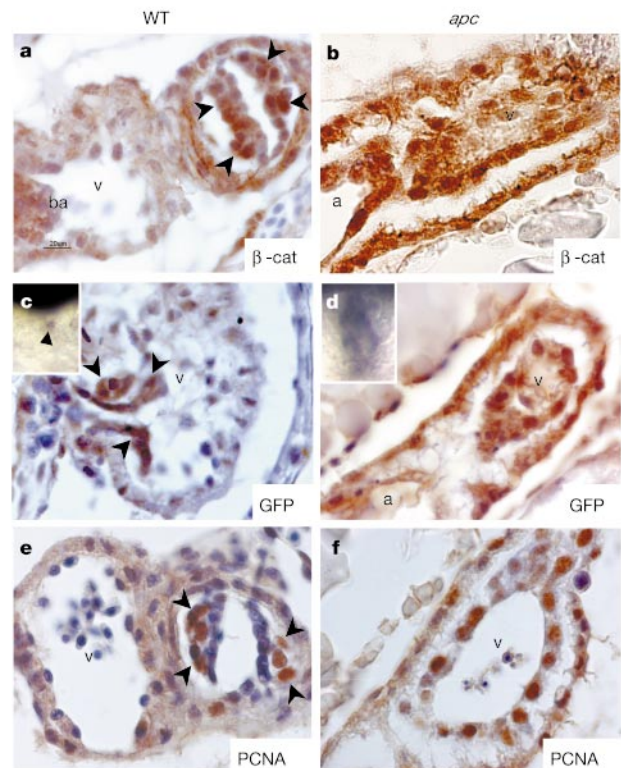


Figure 3 Deregulated Wnt/ β -catenin signalling and proliferation in *apc* mutant hearts. Sagittal sections of WT (a, c, e) and *apc* (b, d, f) hearts at 72 h.p.f. stained for β -catenin (β -cat; a, b), GFP (c, d) and PCNA (e, f). Arrowheads (a, c, e) point to positive nuclei (brown precipitate) of transdifferentiated endocardial cells populating AV cushions. Insets in c and d show whole-mount ISH for GFP expression detected within a few cells only at the AV boundary (arrowhead) in WT; TOPdGFP heart, whereas it is present throughout the anteroposterior extent of the heart in *apc*; TOPdGFP. a, atrium; ba, bulbous arteriosus; v, ventricle. Original magnification $\times 1,000$.

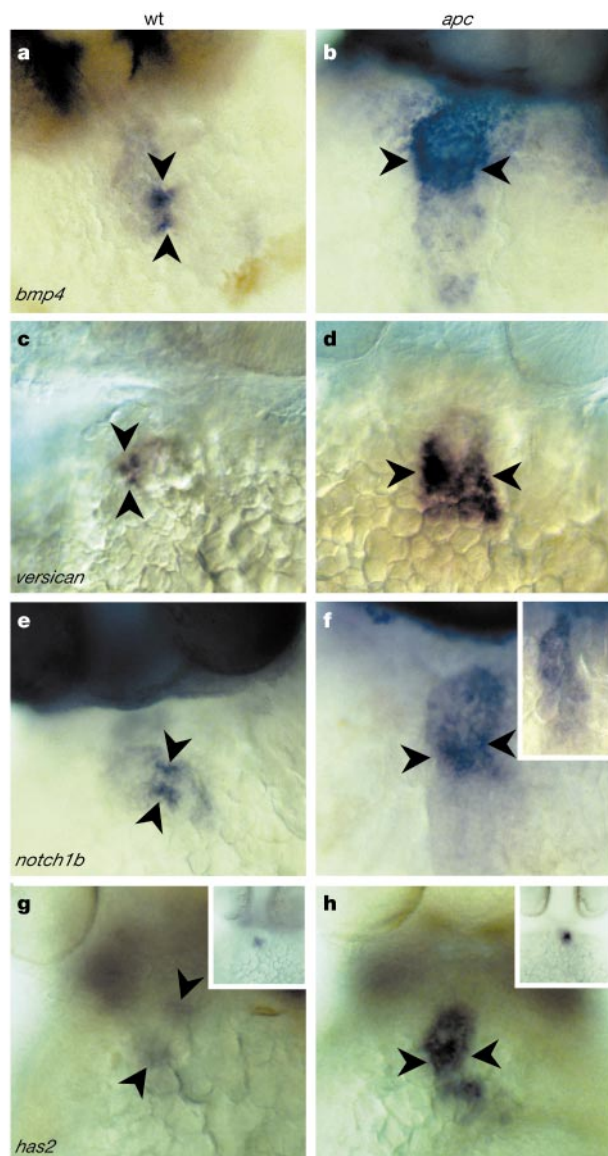


Figure 4 Expression of valve markers is upregulated and expanded in *apc* hearts. Whole-mount ISH of WT (**a, c, e, g**) and *apc* (**b, d, f, h**) embryos at 48 h.p.f. (**a–f**, insets of **g** and **h**) and 72 h.p.f. (**g, h**, inset of **f**). Myocardial *bmp4* and *versican* expression and endocardial *notch1b* and *has2* expression are restricted to the valve-forming region in WT hearts (arrowheads in **a, c, e, g**), whereas expression of all genes is upregulated and domains expanded throughout the *apc* hearts (arrowheads in **b, d, f, h**). Inset in **f** shows *notch1* expression at 72 h.p.f. Insets in **g, h** show *has2* expression being already upregulated at 48 h.p.f. in *apc* hearts. Original magnification $\times 125$.

mutant zebrafish lack functional AV valve tissue. The gene mutated in *jekyll* mutants encodes uridine 5'-diphosphate (UDP)-glucose dehydrogenase¹⁰. *Sugarless*, the *jekyll* orthologue in *Drosophila*, is required for Wnt signalling¹⁸. Wnt/ β -catenin signalling is therefore probably impaired in *jekyll* mutants. Previously, Wnt/ β -catenin signalling has been shown to antagonize cardiogenesis by suppressing cardiomyocyte precursor differentiation from mesoderm¹⁹, demonstrated by conditional deletion of the β -catenin gene, which leads to the formation of multiple ectopic hearts²⁰. Wnt/ β -catenin is also implicated in the proliferation and migration of neural crest cells required for outflow tract septation²¹. Our results now show a distinct role for this pathway in controlling endocardial cell fate decisions and proliferation, thereby modelling the heart proper.

Methods

Target-selected inactivation of the *apc* gene

Fish embryos were raised and staged as previously described²². A sequence contig for the 3'-half of the *apc* gene was constructed by performing database searches in the zebrafish trace database (http://www.ensembl.org/Danio_rerio). We modified a previously described methodology for reverse genetics in zebrafish³ by incorporating *Cell* mediated mismatch recognition²³. Primers (Supplementary Table 1) were designed for the amplification of three overlapping PCR (polymerase chain reaction) fragments (1 kilobase each) and used to screen a library of 4,608 ENU-mutagenized F₁ fish. An F₁ founder harbouring the *apc*^{mc} mutation was outcrossed to Ab and TL wild-type backgrounds, yielding identical phenotypes. To identify the *CA50a* mutation, *apc* complementary DNA generated by reverse transcription coupled PCR (primer sequences available on request) on 60 pooled mutant embryos or 60 wild-type siblings was sequenced.

RNA injections

5'-capped mRNAs were synthesized from pCS2+ constructs encoding APC-GFP²⁴ and zebrafish Dkk1⁶ using mMessage mMachine *in vitro* transcription kit (Ambion). RNA in water with 0.2% phenol red was injected into 1–2-cell-stage embryos of Ab or TL strain.

Histology, ISH and immunohistochemistry

Whole-mount ISH was performed as previously described²². Probes for *bmp4* and *notch1b* have been previously described¹⁰. *has2* and *versican* probes were obtained from J. Bakkers. For (immuno-)histochemistry, dehydrated embryos were paraffin-embedded and sectioned at 6 μ m. Methylene blue was used for routine histology. Immunohistochemical staining with antibodies for β -catenin (Transduction Laboratories) and PCNA (PC10; Euro Diagnostica) were as previously described⁴. GFP protein was detected using antibody B-2 (Santa Cruz Biotechnology). Serial sections were mounted on the same slide, allowing direct comparison of sections. Visualization was with HRP and DAB.

Received 7 July; accepted 18 August 2003; doi:10.1038/nature02028.

1. Fodde, R., Smits, R. & Clevers, H. APC, signal transduction and genetic instability in colorectal cancer. *Nature Rev. Cancer* **1**, 55–67 (2001).
2. Fodde, R. *et al.* A targeted chain-termination mutation in the mouse *Apc* gene results in multiple intestinal tumors. *Proc. Natl Acad. Sci. USA* **91**, 8969–8973 (1994).
3. Mao, B. *et al.* LDL-receptor-related protein 6 is a receptor for Dickkopf proteins. *Nature* **411**, 321–325 (2001).
4. van de Wetering, M. *et al.* The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* **111**, 241–250 (2002).
5. Wienholds, E., Schulte-Merker, S., Walderich, B. & Plasterk, R. H. Target-selected inactivation of the zebrafish *ragl1* gene. *Science* **297**, 99–102 (2002).
6. Hashimoto, H. *et al.* Zebrafish Dkk1 functions in forebrain specification and axial mesendoderm formation. *Dev. Biol.* **217**, 138–152 (2000).
7. Dorsky, R. L., Sheldahl, L. C. & Moon, R. T. A transgenic *Lef1*/beta-catenin-dependent reporter is expressed in spatially restricted domains throughout zebrafish development. *Dev. Biol.* **241**, 229–237 (2002).
8. Heisenberg, C. P. *et al.* A mutation in the Gsk3-binding domain of zebrafish Masterblind/Axin1 leads to a fate transformation of telencephalon and eyes to diencephalon. *Genes Dev.* **15**, 1427–1434 (2001).
9. van de Water, S. *et al.* Ectopic Wnt signal determines the eyeless phenotype of zebrafish masterblind mutant. *Development* **128**, 3877–3888 (2001).
10. Walsh, E. C. & Stainier, D. Y. UDP-glucose dehydrogenase required for cardiac valve formation in zebrafish. *Science* **293**, 1670–1673 (2001).
11. Kielman, M. F. *et al.* *Apc* modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. *Nature Genet.* **32**, 594–605 (2002).
12. Willert, J., Eppig, M., Pollack, J. R., Brown, P. O. & Nusse, R. A transcriptional response to Wnt protein in human embryonic carcinoma cells. *BMC Dev. Biol.* **2**, 8 (2002).
13. Brown, C. B., Boyer, A. S., Runyan, R. B. & Barnett, J. V. Requirement of type III TGF-beta receptor for endocardial cell transformation in the heart. *Science* **283**, 2080–2082 (1999).
14. Kim, R. Y., Robertson, E. J. & Solloway, M. J. Bmp6 and Bmp7 are required for cushion formation and septation in the developing mouse heart. *Dev. Biol.* **235**, 449–466 (2001).
15. Iwamoto, R. Heparin-binding EGF-like growth factor and ErbB signaling is essential for heart function. *Proc. Natl Acad. Sci. USA* (2003).
16. Mjaatvedt, C. H., Yamamura, H., Capehart, A. A., Turner, D. & Markwald, R. R. The *Cspg2* gene, disrupted in the hdf mutant, is required for right cardiac chamber and endocardial cushion formation. *Dev. Biol.* **202**, 56–66 (1998).
17. Camenisch, T. D. *et al.* Disruption of hyaluronan synthase-2 abrogates normal cardiac morphogenesis and hyaluronan-mediated transformation of epithelium to mesenchyme. *J. Clin. Invest.* **106**, 349–360 (2000).
18. Hacker, U., Lin, X. & Perrimon, N. The *Drosophila* sugarless gene modulates Wingless signaling and encodes an enzyme involved in polysaccharide biosynthesis. *Development* **124**, 3565–3573 (1997).
19. Tzahor, E. & Lassar, A. B. Wnt signals from the neural tube block ectopic cardiogenesis. *Genes Dev.* **15**, 255–260 (2001).
20. Lickert, H. *et al.* Formation of multiple hearts in mice following deletion of beta-catenin in the embryonic endoderm. *Dev. Cell* **3**, 171–181 (2002).
21. Kioiusi, C. *et al.* Identification of a Wnt/Dvl/beta-Catenin–Pitx2 pathway mediating cell-type-specific proliferation during development. *Cell* **111**, 673–685 (2002).
22. Westerfield, M. *The Zebrafish Book* (Univ. Oregon Press, Salem, Oregon, 1995).
23. Colbert, T. *et al.* High-throughput screening for induced point mutations. *Plant Physiol.* **126**, 480–484 (2001).
24. Miller, J. R. & Moon, R. T. Analysis of the signaling activities of localization mutants of beta-catenin during axis specification in *Xenopus*. *J. Cell Biol.* **139**, 229–243 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Dorsky and R. Moon for TOPdGFP fish and the APC-GFP construct; M. Kusters and J. Mudde for library screening; and J. Bakkers, M. Morkel and W. Birchmeier for sharing reagents and observations before publication.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.C. (clevers@niob.knaw.nl).

A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals

Esben Bjørn Madsen¹, Lene Heegaard Madsen¹, Simona Radutoiu¹, Magdalena Olbryt¹, Magdalena Rakwalska¹, Krzysztof Szczygłowski², Shusei Sato³, Takakazu Kaneko³, Satoshi Tabata³, Niels Sandal¹ & Jens Stougaard¹

¹Laboratory of Gene Expression, Department of Molecular Biology, University of Aarhus, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark

²Agriculture and Agri-Food Canada, SCPFRC, 1391 Sandford Street, London, Ontario NV5 4T3, Canada

³Kazusa DNA Research Institute, Kisarazu, Chiba, 292-0812, Japan

Plants belonging to the legume family develop nitrogen-fixing root nodules in symbiosis with bacteria commonly known as rhizobia. The legume host encodes all of the functions necessary to build the specialized symbiotic organ, the nodule, but the process is elicited by the bacteria^{1–3}. Molecular communication initiates the interaction, and signals, usually flavones, secreted by the legume root induce the bacteria to produce a lipochitin-oligosaccharide signal molecule (Nod-factor), which in turn triggers the plant organogenic process^{4–7}. An important determinant of bacterial host specificity is the structure of the Nod-factor, suggesting that a plant receptor is involved in signal perception and signal transduction initiating the plant developmental response^{8,9}. Here we describe the cloning of a putative Nod-factor receptor kinase gene (*NFR5*) from *Lotus japonicus*. *NFR5* is essential for Nod-factor perception and encodes an unusual transmembrane serine/threonine receptor-like kinase required for the earliest detectable plant responses to bacteria and Nod-factor. The extracellular domain of the putative receptor has three modules with similarity to LysM domains known from peptidoglycan-binding proteins and chitinases. Together with an atypical kinase domain structure this characterizes an unusual receptor-like kinase.

Inactivation of receptor genes typically results in a phenotype where mutants are unresponsive towards the signal normally perceived by the receptor. In *Lotus nfr5* mutants are non-nodulating and are unresponsive to inoculation with *Mesorhizobium loti* or application of purified bacterial Nod-factor signal molecules. Root hair deformation and the early physiological changes observed in wild-type plants shortly after application of Nod-factor are undetectable in *nfr5* mutants¹⁰. In contrast, mycorrhizal symbiosis with the fungus *Glomus intraradices* is normal in the mutants¹¹, indicating that *NFR5* acts upstream of the common pathway shared between the fungal and bacterial endosymbiotic systems¹². Together these phenotypic characteristics suggest that *NFR5* is required for perception of the Nod-factor signal and subsequent rhizobia-specific activation of the common pathway.

In order to identify and characterize this putative Nod-factor receptor we initiated map-based cloning of the *NFR5* gene. On the

genetic map the *NFR5* locus (formerly known as *SYM5*) was positioned to the lower arm of *Lotus* chromosome II between the AFLP markers E33M40-21 and E32M44-13c (ref. 13). The positional cloning strategy for *NFR5* and the physical map is outlined in Fig. 1a–c and described in the Methods. A contig of TAC and BAC clones was assembled using closely linked markers and sequenced as part of the *Lotus* genome-sequencing programme¹⁴. Subsequent fine mapping located *NFR5* to a 150-kilobase (kb) region delimited by recombination events (Fig. 1b, c). Considering the mutant phenotype, two putative transmembrane receptor kinase genes present among 13 genes in the sequenced region were considered as candidate genes. Sequencing of the two receptor kinase genes in the three *nfr5* alleles identified mutations in one of the genes. We identified an in-frame deletion removing 27 nucleotides in *nfr5-1*, a retrotransposon insertion in *nfr5-2*, and a point mutation leading to

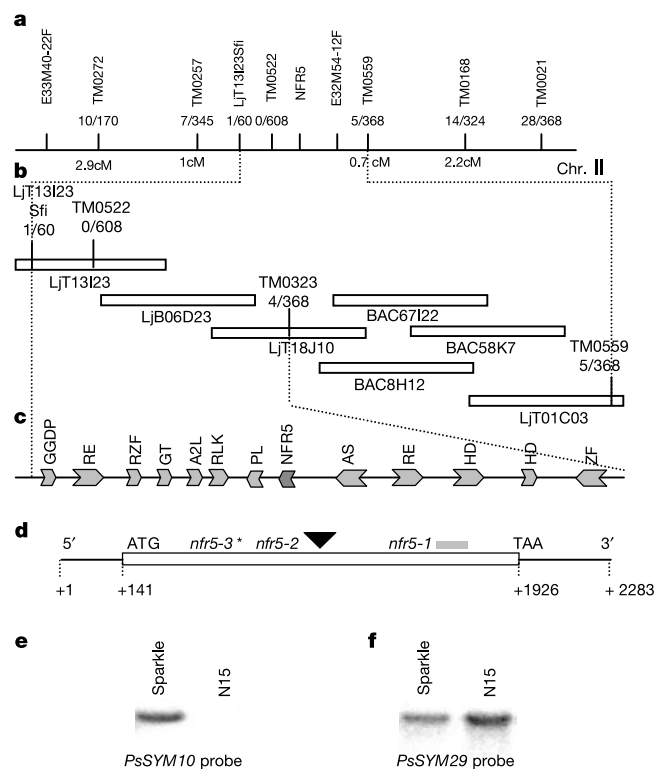


Figure 1 Map-based cloning of *NFR5*. **a**, Genetic map of the *NFR5* region with positions of linked AFLP and microsatellite markers above the line and distances in cM below. The fraction of recombinant plants detected in the mapping population is indicated. **b**, Physical map of the BAC and TAC clones between the closest linked microsatellite markers. The position of sequence-derived markers used to fine map the *NFR5* locus and the fraction of recombinant plants found in the mapping population are indicated. **c**, Candidate genes identified in the sequenced region delimited by the closest linked recombination events. GGDP, geranylgeranyl diphosphate synthase; RE, retroelement; RZF, ring zinc finger protein; GT, glycosyl transferase; A2L, apetal2-like protein; RLK, receptor-like kinase; PL, pectate lyase-like protein; AS, ATPase-subunit; HD, homeodomain protein; ZF, zinc finger protein. Unlabelled, hypothetical proteins. **d**, Structure of the *NFR5* gene, position of the transcription initiation point and the *nfr5-1*, *nfr5-2* and *nfr5-3* mutations. Asterisk, stop codon in *nfr5-3*; black triangle, retrotransposon insertion in *nfr5-2*; grey box, indicates the deletion in *nfr5-1*. **e**, Southern hybridization demonstrating deletion of *SYM10* in the N15 *sym10* mutant line. *EcoRI*-digested genomic DNA of the parental variety Sparkle and the fast-neutron-derived mutant N15 hybridized with a pea *SYM10* probe covering the region encoding the predicted extracellular domain. Lack of hybridization with a probe from the 3' UTR confirmed that the gene was deleted (not shown). **f**, Control hybridization of the same Southern filter using a probe detecting the *P. sativum* *SYM29* gene³¹. The parental *EcoRI* bands are approximately 5 and 9 kb.

a premature stop codon in *nfr5-3* (Figs 1d and 2b). Physical and genetic map location and the identification of mutations in three independent alleles unequivocally identify the receptor kinase gene as *NFR5*. *In planta* complementation of all three *nfr5* mutant alleles obtained with an *Agrobacterium rhizogenes* 'hairy root' transformation system confirms this conclusion, and demonstrates root

control of the non-nodulation phenotype (Table 1). A full-length complementary DNA corresponding to the *NFR5* receptor kinase gene was isolated using a combination of 5' and 3' rapid amplification of cDNA ends (RACE). Alignment of the cDNA with the genomic sequence defined a gene structure without any introns as outlined in Fig. 1d. Sequencing of 18 independent 5' RACE products determined the transcription initiation point and a full-length cDNA of 2,283 nucleotides. An open reading frame of 1,785 nucleotides is preceded by a leader sequence of 140 nucleotides and the 3' untranslated region (UTR) is 358 nucleotides.

The conceptual *NFR5* protein consists of 596 amino acids with a predicted mass of 65.3 kDa (Fig. 2a, b). At the amino terminus, a hydrophobic stretch of 26 amino acids is predicted to act as a signal peptide (Fig. 2a, b). The signal peptide is followed by three motifs with similarity to the LysM domain¹⁵ found, for example, in bacterial peptidoglycan-binding proteins¹⁵ and chitinases from yeast¹⁶ (*Kluyveromyces lactis*) and alga¹⁷ (*Volvox carteri*) (Fig. 2c). The presumed intracellular kinase domain of *NFR5* has motifs associated with functional serine/threonine kinases, except that motifs VII and VIII may be strongly modified or absent. An aspartic acid residue^{18,19} conserved in domain VII is missing and domain VIII comprising the activation loop is highly diverged or is absent^{18,19} (Fig. 2d). Between the three LysM domains and the kinase domain, the algorithms predict a membrane-spanning region. From this analysis a topology with three extracellular LysM receiver domains and an intracellular kinase domain is suggested. The stop codon in *nfr5-3* and the retrotransposon insertion in *nfr5-2* are predicted to generate truncated translation products of 54 and 233 amino acids lacking the LysM motifs and entire carboxy-terminal kinase or lacking just the entire C-terminal kinase, respectively. In the *nfr5-1* allele a 27-nucleotide deletion removes 9 amino acids around kinase domain V, presumably affecting either protein structure or activity. Comparative analysis defines *NFR5* as a founding member of a new family of transmembrane receptor-like kinases including, so far, one from pea, *Medicago truncatula* and rice (Fig. 2c; see also Supplementary Information). Similar to the *NFR5* protein, these putative serine/threonine receptor kinases from pea, *Medicago* and rice lack the aspartic acid residue in domain VII, and the activation loop in domain VIII is highly diverged or absent (Fig. 2d). *Arabidopsis* has five uncharacterized LysM-type kinases. Two of them are intron-less like *NFR5* but none lack the activation loop. With respective amino acid sequence identities of 35%, 42%, 76% and 74%, *Arabidopsis* receptor kinase gene At2g33580, rice Ac103891, *M. truncatula* Ac126779 and pea *SYM10* (see below) encode the proteins most similar to *NFR5*.

The non-nodulating pea *sym10* mutants^{20,21} that display normal mycorrhization but lack root hair deformation in response to rhizobia is the mutant class most similar to *Lotus nfr5*. Macro-synteny encompassing the *SYMRK*, *SHMT*, *ENOD40* and *CHS* markers flanking *NFR5* on chromosome II of *Lotus* and the pea *SYM19* (orthologue of *SYMRK*), *SHMT*, *ENOD40* and *CHS* markers flanking *SYM10* on chromosome I of pea^{13,22,23} also supports *SYM10* as a probable candidate for the pea *NFR5* orthologue. To test this hypothesis, the pea homologue of *NFR5* was cloned from the

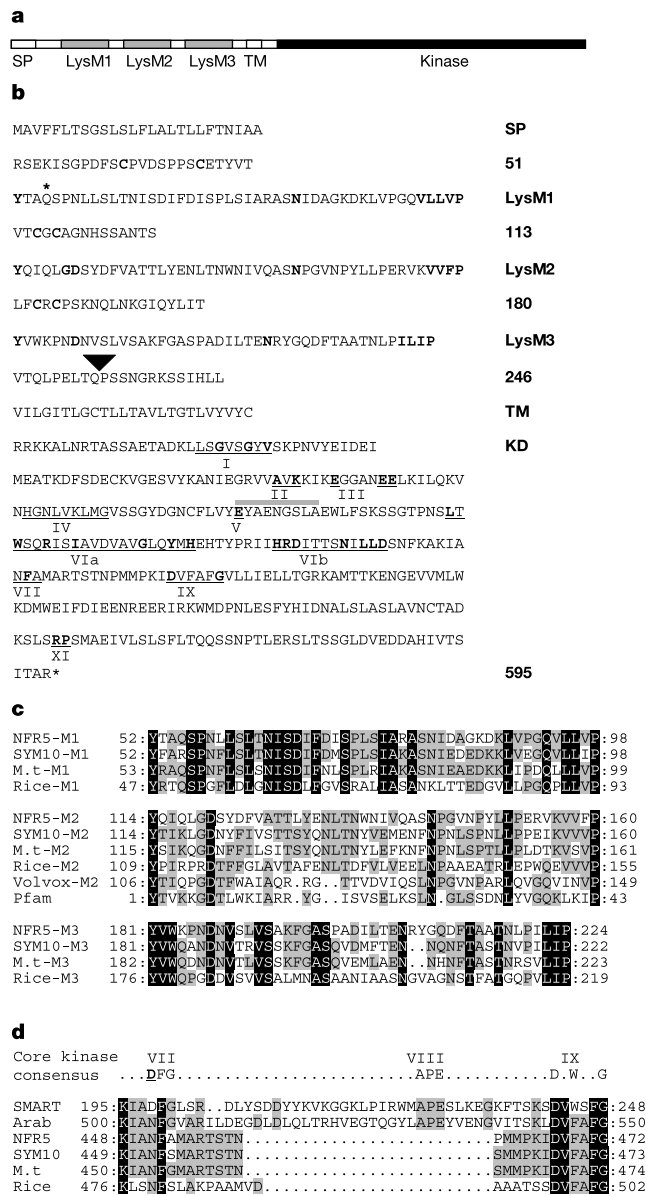


Figure 2 Structure and domains of the *NFR5* protein. **a**, Schematic representation of the *NFR5* protein domains. **b**, Amino acid sequence of *NFR5* arranged in protein domains. Bold indicates conserved LysM residues, whereas bold and underlined residues indicate residues conserved in protein kinase domains (KD)^{18,19}. TM, transmembrane; SP, signal peptide; asterisks, stop codon in *nfr5-3*; black triangle, retrotransposon insertion in *nfr5-2*; grey box, indicates amino acids deleted in *nfr5-1*. **c**, Individual alignment of the three LysM motifs (M1, M2, M3) from *NFR5*, *SYM10*, *M. truncatula* (Ac126779) and rice (AAM19130.1), the single LysM from *V. carteri* (T08150) and the Pfam consensus. **d**, The highly diverse or absent activation loop (domain VIII) in the *NFR5* family of receptor kinases is illustrated by alignment of kinase domains VII, VIII and IX from *Arabidopsis* (NP_180916), *NFR5*, *SYM10*, *M. truncatula* (*M. t.*; Ac126779), rice (AAM19130.1) and the SMART consensus. The aspartic acid normally conserved in domain VII marked in bold and underlined. In **c** and **d** the amino acids conserved in all domains compared are shaded black and those conserved in two or more domains are shaded grey.

Table 1 Complementation of *nfr5-1*, *nfr5-2* and *nfr5-3* by the wild-type *NFR5* gene

Plant genotype	Construct	Fraction of nodulated transgenic roots	Nodules per plant
WT	Vector	19 of 23	3.4
<i>nfr5-1</i>	<i>NFR5</i>	52 of 66	4.5
<i>nfr5-1</i>	Vector	0 of 36	0
<i>nfr5-2</i>	<i>NFR5</i>	11 of 15	7.1
<i>nfr5-2</i>	Vector	0 of 6	0
<i>nfr5-3</i>	<i>NFR5</i>	24 of 32	4
<i>nfr5-3</i>	Vector	0 of 29	0

parental wild-type cultivars 'Sparkle', 'Finale' and 'Frisson' using polymerase chain reaction (PCR) and library screening. Southern blot analysis of N15, the original *sym10* mutant type strain²¹, isolated after fast neutron mutagenesis showed that the *SYM10* gene was deleted (Fig. 1e, f), and sequencing of three independent *sym10* mutant lines identified three mutant alleles of the gene, all carrying nonsense mutations (Table 2). The pea *SYM10* gene encodes a putative serine/threonine receptor-like kinase, *SYM10*, with a predicted molecular mass of 66 kDa. Comparison of the conceptual *NFR5* and *SYM10* proteins reveals an amino acid identity of 74% and overall conservation of the three LysM motifs as well as the kinase domains (Fig. 2; see also Supplementary Information). The premature stop codons identified in pea mutants P5, RisFixG and P56 (refs 20, 24) are predicted to generate truncated *SYM10* proteins of 199, 387 and 404 amino acids, respectively, which lack part of, or the entire kinase domain (Table 1).

As *NFR5* and *SYM10* are potential components of a receptor complex perceiving the rhizobial Nod-factor signal at the root surface, we investigated their expression in various organs of *Lotus* and pea. Northern hybridization demonstrated that expression of *NFR5* is highest in *Lotus* root tissue and has a barely detectable expression in nodules (Fig. 3a). Comparing messenger RNA levels in uninoculated roots with the corresponding transcript levels in *M. loti*-inoculated roots, nodules, leaves, flowers and pods determined an expression level approximately 60- and 120-fold lower in nodules and shoot tissues (Fig. 3b). Northern blot analysis in pea detects transcripts in uninoculated root and a higher level in nodules (Fig. 3a).

Here we describe a component of the receptor that mediates lipochitin-oligosaccharide-induced activation of legume plants for rhizobial invasion and nodule meristem formation. Analysis of legume symbiotic mutants has identified a common gene set required for both mycorrhizal and rhizobial symbiosis. The role of the *NFR5* receptor-like kinase is probably to integrate the rhizobial-induced plant response into the evolutionarily older fungal-induced signal transduction pathway²³. Our data support a hypothesis where the *NFR5* receptor kinase is one of the specialized genes allowing legumes to establish endosymbiosis with nitrogen-fixing bacteria through activation of the common pathway. In this perspective, transfer of the *NFR5* gene to non-legumes may be one of the steps required to establish interaction with nitrogen-fixing rhizobial bacteria. Surprisingly, the monocotyledonous rice has a gene that is more similar to *NFR5* than genes found in the dicotyledonous *Arabidopsis*. Monocotyledonous plants such as the cereals diverged from the eudicot lineage around 150 million years ago²⁵, and the unexpected better conservation of a *NFR5* homologue in a more distant lineage predicts selection for function. One could speculate that rice has maintained an ancient microbial/bacterial interaction that was recruited by legumes for the rhizobial interaction, whereas it has been lost or diverged further in other dicotyledonous plants such as *Arabidopsis*.

Phenotypic and genetic comparison of *Lotus NFR5* and the pea *SYM10* gene demonstrates functional conservation in two legumes that follow the determinate versus indeterminate organogenic

pathways for root nodule formation, respectively. A high expression level of the pea *SYM10* gene in root nodules may reflect the different developmental pathways. Indeterminate pea nodules maintain both the meristematic activity and a zone of invasion in the mature nodule, and thus might require continued *SYM10* function(s). Invasion and meristematic activity within the determinate *Lotus* root nodules terminates early during the organogenic process, probably lowering the requirement for *NFR5* at later stages of nodule development. *Lotus* and pea also represent two different cross-inoculation groups, reflecting the host specificity in legume rhizobial interaction. *Lotus* develops nodules with *M. loti* whereas pea develops nodules after infection with *Rhizobium leguminosarum* biovar *viciae*. These two rhizobial strains secrete structurally different Nod-factors that may be ligands for the *NFR5* and *SYM10* receptor kinases, respectively. Protein domain structure and topology predicts that the extracellular LysM domains, directly or indirectly, would be involved in binding of Nod-factor, thus initiating signal transduction through the intracellular kinase. Detailed comparison of *Lotus* and pea LysM domains of the two extracellular regions show amino acid identities of 74%, 60% and 66% in M1, M2 and M3, respectively. An NMR structure for the *Escherichia coli* MltD LysM domain²⁶ can now serve as a backbone for detailed studies of the role of these motifs in Nod-factor binding. *In vitro* biochemical binding and affinity studies will be necessary to resolve the question of Nod-factor interaction/binding at one or more of the extracellular *NFR5* and *SYM10* LysM domains.

NFR5 represents an unusual family of receptor kinases with an apparent absence of an activation loop. Normally, phosphorylation of the activation loop makes the active site of the kinase available for substrates, thus activating the protein. Lack of an activation loop may suggest that the *NFR5*-type kinases are active once a ligand binds at the extracellular domain. Studies on the interaction between the *NFR5*/*NFR1*-mediated rhizobial signalling¹⁰ and the mycorrhizal- and rhizobial-activated *SYMRK*²³ signal transduction cascades are one of the challenges for the future. □

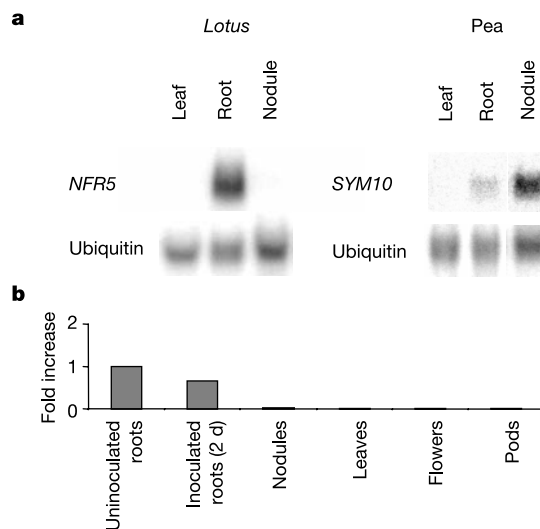


Figure 3 Expression of *Lotus NFR5* and pea *SYM10*. **a**, Northern analysis of *NFR5* and *SYM10* expression in leaf, uninoculated root and nodule of *Lotus* and pea, respectively. Ubiquitin is the loading control. **b**, Real-time PCR with reverse transcription determination of expression levels of *NFR5* in uninoculated roots, *M. loti*-inoculated roots, nodules, leaves, flowers and pods of *Lotus* plants. Results are shown as fold increase compared with uninoculated roots at time zero. The identity of the amplified sequences was confirmed by sequencing. ATPase was used as internal constitutive control, and relative normalized values are shown.

Table 2 Summary of *Lotus nfr5* and pea *sym10* mutant alleles

Allele	Mutation	Species
<i>nfr5-1</i>	EYAENGSLA 380–388 deletion	<i>Lj</i>
<i>nfr5-2</i>	Retrotransposon integration after Q233	<i>Lj</i>
<i>nfr5-3</i>	CAG → TAG, Q55 → stop	<i>Lj</i>
RisFixG	TGG → TGA, W388 → stop	<i>Ps</i>
P5	TGG → TGA, W405 → stop	<i>Ps</i>
P56	CAA → TAA, Q200 → stop	<i>Ps</i>
N15	<i>sym10</i> gene deleted	<i>Ps</i>

Lj, *L. japonicus*; *Ps*, *P. sativum*.

Methods

Plant material

The *nfr5-1*, *nfr5-2* (previously called *sym5*) and *nfr5-3* (previously called *sym25*) mutants were identified in a screen for T-DNA tagging, a screen for mutations induced by the transposable element *Ac*, and an ethylmethane-sulphonate-mutagenized population, respectively^{27,28}. Plants were grown with or without *M. loti* NZP2235 as previously outlined²⁹. The pea *sym10* complementation group has been described previously²². Pea mutants were obtained from K. Engvild (Riso National Laboratory) and from the pea germplasm collection at JIC Norwich, UK (<http://www.jic.bbsrc.ac.uk/germplas/pisum/index.htm>).

Mapping population

Intra- and interspecific F₂ mapping populations were established by crossing a *L. japonicus* 'Gifu' *nfr5-1* mutant with wild-type *L. japonicus* ecotype 'MG20' and wild-type *L. filicaulis*¹³. F₂ plants homozygous for the *nfr5-1* mutant allele were identified after screening for the non-nodulation mutant phenotype. A total of 240 homozygous F₂ mutant plants were analysed in the *L. filicaulis* mapping population and 368 homozygous F₂ mutant plants in the MG20 mapping population.

Positional cloning and complementation

AFLP and bulked segregant analyses were performed using the *EcoRI*/*MseI* restriction enzyme combination¹³. Initially, *NFR5* was mapped between AFLP markers E33M40-22F and E32M54-12F with the *L. filicaulis*-based mapping population. The E32M54-12F marker was cloned and used to isolate BAC clones 8H12 and 67I22, and TAC clone LJ18J10. The ends of the contig formed by these three clones were used to isolate new BAC and TAC clones—on one side BAC clone 58K7 and TAC clone LJ101C03 and on the other side BAC clone LJ106D23. The outer end of LJ106D23 was used to isolate TAC clone LJ13123. Sequence-anchored markers from this contig were mapped, and in the *L. filicaulis* mapping population one recombinant plant was found with a marker (LJ13123Sf) on the outer end of LJ13123, whereas no recombinant plants were found with a marker (TM0522) from the middle of this TAC. In the *L. japonicus* MG20 mapping population 4 out of 368 recombinant plants were found with marker TM0323, delimiting *NFR5* to a region of 150 kb containing 13 genes.

For *in planta* complementation a 3.5-kb fragment containing the *NFR5* gene plus a 1,175-base pair (bp) promoter region and a 441-bp 3' UTR was used. After insertion into pIV10, tri-parental mating was performed to introduce the plasmid into *A. rhizogenes* strain AR12 and AR1193. A transgenic hairy roots and nodulation test was done as previously described³⁰.

cDNA isolation

RNA extracted from roots grown without nitrate and rhizobia was used to make a 5' RACE-ready cDNA pool using the SMART RACE cDNA amplification kit (Clontech). 5' RACE was performed with a 3' UTR primer 5'-GCTAGTTAAAATGTAATAGTAAACCA CGC-3' and the SMART oligonucleotide provided in the kit. The 5' RACE product of approximately 2 kb was cloned in a TOPO vector (Invitrogen). Eighteen clones were sequenced with the longest having a 140-bp 5' leader region. On the same 5' RACE-ready cDNA a modified 3' RACE was performed. A gene-specific primer 5'-AAAGCAGCATT CATCTTCTGG-3' and an oligo-dT primer was used for five rounds of PCR at 42 °C and then for 30 cycles at 58 °C. The PCR products were used as template for a new PCR reaction with a nested gene-specific primer 5'-GCAAGGGAAGGTAATTCAG-3' and the same oligo-dT primer. This PCR product was cloned and three independent clones were sequenced.

Cloning of the pea *SYM10* gene

The forward primer 5'-ATGCTGCTTCTTCTTCTCCTTC-3' and the reverse primer 5'-CCACACATAAGTAATMAGATACT-3' for amplification of a pea *SYM10* fragment were designed based on conserved regions of the *L. japonicus* *NFR5* and the *M. truncatula* root expressed sequence tag homologue BE204912. The 551-bp cv. Finale product was cloned, sequenced and used as a hybridization probe to isolate a genomic clone (cv. Alaska) and a full-length cDNA clone (cv. Finale). On the basis of the sequences of these clones the sequences of the gene in cv. Frisson and cv. Sparkle were determined by a PCR amplification/sequencing approach.

Southern hybridization and expression analysis

The probes used for Southern and northern analyses are listed below. *NFR5* probe, 790-bp PCR fragment produced by forward primer 5'-AGTTCCTTTTCTCTTCCCTG-3' and reverse primer 5'-CCAGAGATGAATGCTGCTTT-3' (52-bp 5' UTR plus the region encoding amino acids 1–248 of the predicted extracellular domain). *SYM10* probe, 602-bp PCR fragment produced by forward primer 5'-CCATAACTGCAATTCCTTAC-3' and reverse primer 5'-GTAACATTGTCATTAGCCTGCC-3' (33-bp 5' UTR plus the region encoding amino acids 1–190 of the predicted extracellular domain). *Lotus japonicus* ubiquitin probe, 230-bp PCR fragment produced by forward primer 5'-ATGCAGATCT TTTGTGAAGAC-3' and reverse primer 5'-ACCACCACGGAAGACGGAG-3' covering a single ubiquitin repeat. Different tissues were collected and the level of *NFR5* steady-state mRNA determined by quantitative real-time PCR and northern hybridization as described¹⁰.

Computer analysis

Sequences were analysed by BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) and PredictProtein (<http://www.embl-heidelberg.de/predictprotein/predictprotein.html>). The signal peptide was predicted by SignalP v.1.1 (<http://www.cbs.dtu.dk/services/>

SignalP) and the transmembrane region was predicted by TMHMM v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). We used SMART and Pfam for domain analysis (<http://www.sanger.ac.uk/Software/Pfam/> and <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Received 16 July; accepted 11 September 2003; doi:10.1038/nature02045.

- Truchet, G. *et al.* Alfalfa nodulation in the absence of *Rhizobium*. *Mol. Gen. Genet.* **219**, 65–68 (1989).
- Truchet, G. *et al.* Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. *Nature* **351**, 670–673 (1991).
- Spaink, H. P. *et al.* A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. *Nature* **354**, 125–130 (1991).
- Lerouge, P. *et al.* Symbiotic host specificity of *Rhizobium meliloti* is determined by a sulphated and acetylated glucosamine oligosaccharide signal. *Nature* **344**, 781–784 (1990).
- Lopez-Lara, I. M. *et al.* Structural identification of the lipo-chitin oligosaccharide nodulation signals of *Rhizobium loti*. *Mol. Microbiol.* **15**, 627–638 (1995).
- Niwa, S. *et al.* Responses of a model legume *Lotus japonicus* to lipochitin oligosaccharide nodulation factors purified from *Mesorhizobium loti* JRL501. *Mol. Plant Microbe Interact.* **14**, 848–856 (2001).
- Long, S. R. *Rhizobium* symbiosis: nod factors in perspective. *Plant Cell* **8**, 1885–1898 (1996).
- Ardourel, M. *et al.* *Rhizobium meliloti* lipooligosaccharide nodulation factors: different structural requirements for bacterial entry into target root hair cells and induction of plant symbiotic developmental responses. *Plant Cell* **6**, 1357–1374 (1994).
- Pacios-Bras, C. *et al.* A *Lotus japonicus* nodulation system based on heterologous expression of the fucosyl transferase *NodZ* and the acetyl transferase *NolL* in *Rhizobium leguminosarum*. *Mol. Plant Microbe Interact.* **13**, 475–479 (2000).
- Radutoiu, S. *et al.* Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**, 585–592 (2003).
- Wegel, E., Schauer, L., Sandal, N., Stougaard, J. & Parniske, M. Mycorrhiza mutants of *Lotus japonicus* define genetically independent steps during symbiotic infection. *Mol. Plant Microbe Interact.* **11**, 933–936 (1998).
- Stougaard, J. Genetics and genomics of root symbiosis. *Curr. Opin. Plant Biol.* **4**, 328–335 (2001).
- Sandal, N. *et al.* A genetic linkage map of the model legume *Lotus japonicus* and strategies for fast mapping of new loci. *Genetics* **161**, 1673–1683 (2002).
- Nakamura, Y. *et al.* Structural analysis of a *Lotus japonicus* genome. II. Sequence features and mapping of sixty-five TAC clones which cover the 6.5-Mb regions of the genome. *DNA Res.* **9**, 63–70 (2002).
- Steen, A. *et al.* Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *J. Biol. Chem.* **278**, 23874–23881 (2003).
- Butler, A. R., O'Donnell, R. W., Martin, V. J., Gooday, G. W. & Stark, M. J. *Kluyveromyces lactis* toxin has an essential chitinase activity. *Eur. J. Biochem.* **199**, 483–488 (1991).
- Amon, P., Haas, E. & Sumper, M. The sex-inducing pheromone and wounding trigger the same set of genes in the multicellular green alga *Volvox*. *Plant Cell* **10**, 781–789 (1998).
- Schenk, P. W. & Snaar-Jagalska, B. E. Signal perception and transduction: the role of protein kinases. *Biochim. Biophys. Acta* **1449**, 1–24 (1999).
- Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).
- Duc, G. & Messager, A. Mutagenesis of pea (*Pisum sativum* L.) and the isolation of mutants for nodulation and nitrogen fixation. *Plant Sci.* **60**, 207–213 (1989).
- Kneen, B. E., Weeden, N. F. & LaRue, T. A. Non-nodulating mutants of *Pisum sativum* (L.) cv. Sparkle. *J. Hered.* **85**, 129–132 (1994).
- Schneider, A. *et al.* Mapping of the nodulation loci *sym9* and *sym10* of pea (*Pisum sativum* L.). *Theor. Appl. Genet.* **104**, 1312–1316 (2002).
- Stracke, S. *et al.* A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* **417**, 959–962 (2002).
- Engvild, K. C. Nodulation and nitrogen fixation mutants of pea, *Pisum sativum*. *Theor. Appl. Genet.* **74**, 711–713 (1987).
- Wikstrom, N., Savolainen, V. & Chase, M. W. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B* **268**, 2211–2220 (2001).
- Bateman, A. & Bycroft, M. The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J. Mol. Biol.* **299**, 1113–1119 (2000).
- Schauer, L. *et al.* Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. *Mol. Gen. Genet.* **259**, 414–423 (1998).
- Szczygłowski, K. *et al.* Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Mol. Plant Microbe Interact.* **11**, 684–697 (1998).
- Handberg, K. & Stougaard, J. *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J.* **2**, 487–496 (1992).
- Stougaard, J. *Agrobacterium rhizogenes* as a vector for transforming higher plants. *Methods Mol. Biol.* **49**, 49–61 (1995).
- Krusell, L. *et al.* Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* **420**, 422–426 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Ambrose, A. Downie and K. Engvild for providing seeds from plants with *sym10* alleles. S.R. was supported by an EU Marie Curie Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S. (stougaard@mb.au.dk). The sequences for the *L. japonicus* ecotype Gifu *NFR5* gene and mRNA are deposited in the EMBL database under accession numbers AJ575254 and AJ575255; the *Pisum sativum* *SYM10* gene and mRNA sequences are deposited under accession numbers AJ575250 cv. Alaska, AJ575251 cv. Frisson, AJ575252 cv. Sparkle and AJ575253 cv. Finale. Sequences of MG20 TAC clones are deposited in GenBank under accession numbers LJ18J10a AP006538, LJ18J10b AP006539, LJ101C03 AP006540, LJ106D23 AP006541 and LJ13123 AP006542.

Take a closer look

Light, stereo, scanning, transmission and other angles on microscopy.

MZ16 and MZ16A

Leica Microsystems

www.leica-microsystems.com

High resolution and a motorized zoom

These stereomicroscopes from Leica Microsystems feature a resolution of 840 line pairs per mm, $\times 7.1$ to $\times 230$ magnification, a motor zoom, automatic measuring functions and a digital display. Structures as small as $0.6\text{ }\mu\text{m}$ can be observed. The motor zoom is controlled using movements of the finger, the feet or via a computer. With each zoom change, the MZ16A immediately displays the current magnification and measurements in mm, inches or thousandths of mm.



M rules: the Leica MZ16A stereomicroscope.

Camedia C-5050

Olympus

www.olympus-europa.com

Fine control for photomicrography

The Camedia C-5050 from Olympus is a 5-megapixel CCD designed for photomicrography recording, archiving, analysis and transmission. It offers manual control over aperture, shutter speed and focus. Sharpness control can also be left to the camera's automatic contrast detection focusing system. The processing engine enables continuous shooting of either 11 images at 1.7 frames per second, or four shots at 3 frames per second. The tilting LCD monitor on the rear of the camera makes it easy to view the image when the camera is mounted on a vertical microscope port. The Camedia is compatible with xD-Picture Card as well as SmartMedia, Compact Flash and Microdrive, and is integrated with DP-SOFT software. It also has a USB autoconnect audio/video interface.

ErgoScan

Prior Scientific

www.prior.com

A hands-down winner

Prior's ErgoScan system has been designed to minimize the ergonomic problems associated with using a microscope. Motorized scanning stage and focusing mechanisms eliminate the need to stretch and twist microscope controls, reducing operator fatigue and discomfort. Direction and speed of scanning are controlled using a low-resistance joystick, while focusing is carried out using a flat hand roller. Both offer fine control and minimal resistance. The device can be added to most microscopes.

H-9500 TEM

Hitachi High-Technologies

www.hitachi-hitec-uk.com

Five-minute wonder

The PC-controlled H-9500 transmission electron microscope can be readied for use in less than five minutes. Specimen changes take a minute and a stable electron beam at 300kV is produced within three minutes. It uses a single crystal LaB₆ electron source and offers 0.102 nm crystal lattice resolution and 0.18 nm point-to-point resolution. Using Windows 2000, up to 20 sets of operating conditions (such as lens settings and accelerating voltage) can be saved for easy recall. The microscope is equipped with a high-resolution CCD camera with fast or slow scan functions for direct recording of digital images. Images are stored on a database and can be viewed as thumbnails. Specimen positions may be viewed simply by recalling specimen stage coordinates.

AxioCam MRc5

Carl Zeiss

www.zeiss.de/micro

See its true colours

The AxioCam MRc5 colour camera has a high-resolution, 5-megapixel CCD sensor for rapid production of true-colour micrographs in applications requiring coloured light microscope specimens or colour contrasting techniques. A FireWire interface allows connection of the camera to a PC. Recording software allows convenient control of the camera, with a live image of the microscope specimen provided to aid handling and setting of the specimen on the microscope.

SynchroScan

Synchroscopy

www.synchroscopy.com

Take control a stage further

SynchroScan, an automated stage and focus control system, is now available with auto-



The ErgoScan kit for ergonomic microscopy.

matic, motorized control of the X-Y stage, Z-axis, illumination, condenser aperture diaphragm and top lens, as well as objective turret control of Olympus BX61 or Leica DMR microscopes. The focus mechanism of a range of Leica stereomicroscopes can also be controlled. For high-speed slide scanning or image mapping users can remotely control the X, Y and Z axes with SynchroScan RT to scan specimens at up to 20 fields per second. SynchroScan can combine with Montage Explorer software to capture high-resolution images of up to $16,000 \times 16,000$ pixels from a wide field of view. It can also link with Synchroscopy's AcQuis software for colour image capture and Auto-Montage for three-dimensional imaging.

MT-2000

Intracel

www.intracel.co.uk

Motorized X-Y translation

The MT-2000 motorized translation system is based on the MT-800 stepper-motor-driven X-Y translation table. The MT-800 controller allows robotic control of movement and provides a serial interface. In the complete system the recording chamber and micromanipulators are mounted on independent columns firmly affixed to the user's antivibration table. The chamber column options specified for the MT-1000 are also available for the MT-2000. Motorized translators are available for use with the Olympus BX-51WI, the Nikon E600FN, the Leica DMLFS and the Zeiss Axioskop 2FS.

Cryo-Trim CTT-90 and CTT-45

Diatome

www.emsdiasum.com

A matter of degree

Diatome's diamond trimming tools for light and electron microscopy can be used to trim blocks at both room and cryo temperatures quickly and easily, without the need for razor

blades or other hand-held or mechanical trimming devices. Trimming is done directly in the ultramicrotome and, with a quick switch, final sectioning can begin. The CTT-90 blade provides perfectly trimmed blocks in which the sides of the sections remain straight; hence, with the 90° trimming tool, the sample remains at the same width during sectioning. The CTT-45 blade produces sectioned sides that are pyramidal in shape, with the sections becoming wider as the block is cut.

Structural and chemical analyser for SEM

Renishaw www.renishaw.com

Elemental analysis at micrometre level

Raman spectroscopy provides chemical, physical and structural information on many samples and has, until recently, usually been built into optical microscopes. Renishaw's analyser can be used to combine the resolving and visualizing power of SEM with molecular spectroscopy to allow unambiguous chemical and structural characterization of a wide range of samples at the micrometre scale under HV, UHV or environmental conditions. Fully retractable optics are inserted between the SEM's objective lens and the sample, positioning an analysis spot with sub-micrometre precision and repeatability while retaining the capacity to view the sample using secondary electron imaging, and analyse it using EDS. The collection optics are also suitable for cathodoluminescence and photoluminescence studies. The analyser can be fitted to most SEM makes and models.

Nano-R

Pacific Nanotechnology

www.pacificnanotech.com

Atomic force microscopy for beginners

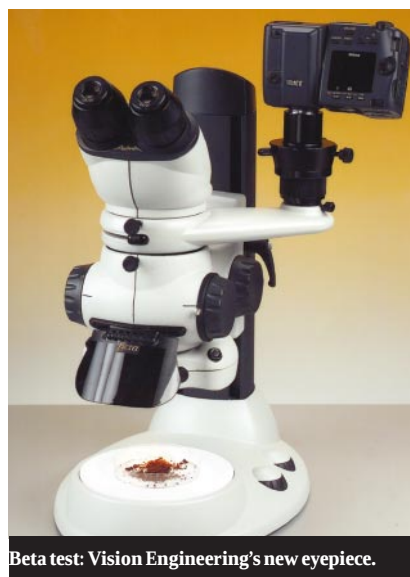
The Nano-R AFM is billed as a powerful but practical and affordable tool that can be used by any researcher, regardless of experience. Novice users are said to be able to obtain images in about an hour. The unit includes SPM Cockpit software for image acquisition and analysis, which can be set either to EZMode for novice users or X'Pert for those who are more experienced. There are two- and three-dimensional display modes and most common analysis techniques, such as line profiles, can be performed. Applications include metallurgy, nanoparticles and powders, filters, semiconductor/data storage, photonics and adhesives.

Beta

Vision Engineering www.visioneng.com

Mid-range view

Beta is a stereo zoom eyepiece microscope for industrial and biomedical applications. It offers parfocal quality zoom from 3× to 160× magnification (standard zoom 6× to 40×)



Beta test: Vision Engineering's new eyepiece.

with semicoaxial spot lighting and substage illumination. It can be mounted on a compact bench stand or a swing-away telescopic boom. Modular accessories include a reference reticle, digital/video camera pick-offs, angled viewing with fixed or rotating angled optical path and fixed or adjustable eyepiece height.

IlluminatIR

SensIR Europe

www.sensir.com

Transforming microscopes into microprobes

The IlluminatIR is a novel system that transforms an infinity-corrected light microscope into a versatile infrared microprobe. The system delivers molecular information while retaining all the functions and performance of the original microscope, including polarized light, fluorescence and image analysis. Users can also continue to make use of the familiar controls of their microscope systems to view, position and manipulate their specimens. To collect spectral information, the user raises the stage to bring the sample into contact with the internal reflection element surface, then presses a button to acquire information and create a report.

AutoIMAGE Version 5.00

PerkinElmer

www.perkinelmer.com

Interactive 3-D mapping

PerkinElmer has released an upgraded version of its AutoIMAGE software for Fourier-transform infrared (FT-IR) microscopy. Version 5.00 offers an extended suite of data-processing algorithms and an enhanced three-dimensional display. Users can employ techniques in which a target spectrum can be matched against or subtracted from every point in a chemical map to enhance small differences in composition. The interactive three-dimensional surface projections allow product defects to be visualized quickly. In addition, chemical maps can be overlaid, enabling the distribution of multiple components across a sample to be seen easily.

Carbon nanotube AFM tips

nPoint

www.piezomax.com

Take a tip from them

nPoint, makers of ultra-precision motion and control nanopositioners for nanoscale research, have made it easier to choose the appropriate tip to image high aspect ratio features for atomic force microscope tips. ProbeMAX carbon nanotube AFM tips now come with an added extra, an SEM image, allowing the user to select the appropriate probe for their application. ProbeMAX carbon nanotube probes consist of multi-wall carbon nanotubes or nanotube bundles mounted on commercially available silicon AFM probes. They offer long lifetime and can replace conventional silicon probe tips in many intermittent-contact mode applications. The tips, with their SEM images, are available as either a single multi-wall nanotube or a bundle of multi-wall nanotubes mounted on silicon probes.

TriPix

Electron Microscopy Sciences

www.emsdiasum.com

Three of a kind

The TriPix digital camera system is available in three models: TriPix-RGB, TriPix-IR and TriPix-Spectral. All offer a resolution of 1,280×1,024 pixels, three-shot colour or instantaneous grayscale image capture and a USB interface. The cameras feature microlenses on each pixel for enhanced sensitivity. Interference filters allow uniform transmission and sharp wavelength transitions for colour fidelity. With three different modes the RGB offers 6 µm square pixel size, while the IR provides 1,280×1,024 pixel imaging in the near IR (700–960 nm) and the Spectral has 1,280×1,024 pixel full-spectrum imaging in user-selectable wavelengths from 400 to 960 nm.

XE-Series

Ambios Technology

www.ambiostech.com

Superior scanning

XE-Series of scanning probe microscopes use a novel X-Y-Z scan system to achieve distortion-free data acquisition. The design incorporates a guided flexure scanner with low inertia, minimal run-out, improved orthogonality and axis-independent performance. It eliminates the cross-talk, non-linearity, creep and hysteresis inherent in the conventional piezoelectric tube scanners used in most scanning probe microscopes. The series includes models to accommodate sample sizes up to 300 mm. Applications include material and surface characterization, semiconductors, polymers, data storage and magnetic materials.

These notes are compiled in the Nature office from information provided by the manufacturers.

Material intended for publication can be sent to newproducts@nature.com

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Global Head of Naturejobs:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)
**Netherlands/ Italy/
Iberia/ Belgium:**
Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)
France/ Switzerland:
Amelie Pequignot (4974)
Natureevents:
Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:
Hind Berrada (4935)
Naturejobs web development:
Tom Hancock
Naturejobs online production:
Ben Lund

European Satellite Office

Germany/ Austria:
Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:
Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:
Rinoko Asami
e-mail: r.asami@naturejpn.com

naturejobs

New perspectives

Why did the DFG, the German Research Foundation, open an office in Washington a year ago? To gain a perspective on how science is conducted in another country, said the DFG's president Ernst-Ludwig Winnacker last week at a reception to welcome the new director of the Washington office, Marina Koch-Krumrei. Naturejobs is opening a new feature this issue, called Career View, for similar reasons. We want to bring in more voices from different career stages — from the postdoc trying to land his or her first tenure-track job to a senior scientist who's bounced between academia and industry.

The first new perspective we've added is that of an outsider. Deb Koen, vice-president of the company Career Development Services based in upstate New York, has years of experience counselling scientists and non-scientists. We know that her 'nuts and bolts' advice, on everything from writing resumés to negotiating salaries, will be valued by scientists, even if they're not actively job-hunting.

Noting the many organizations catering specifically to young scientists — from students to postdocs — popping up around the globe, we thought they deserved a platform. The first Career View features the results of a survey of young biotechnologists by the Association of Italian Biotechnologists. We will try to give a voice to organizations from all over the world. Finally, we have moved our established feature 'Movers' to this new page in an altered format that focuses on one scientist and their career path.

And like the incoming head of DFG's Washington office, Naturejobs also has an interest in addressing so-called 'soft issues', such as finding jobs for spouses when a scientist relocates, or the effect that housing costs have on recruiting postdocs to expensive cities. We encourage suggestions for new issues and angles and welcome the addition of your voice to our Career View.

Paul Smaglik

Naturejobs editor



Contents

CAREER VIEW

Nuts & Bolts

Scientist, know thyself

Young scientist

Italian biotechnologists

Movers

Mark Boguski

p644

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

YOUNG SCIENTIST

Italian biotechnologists organize

Francesco Lescai and Marco Quarta

Current poor prospects for young Italian life scientists are highlighted by a survey by the Association of Italian Biotechnologists (ANBI), released in July 2003. Although Italy has invested in growing its scientific and technical community, jobs have not followed. Young biotechnologists' salaries are low and they are eager to leave Italy.

A sample of 120 young biotechnologists, most of whom have the *laurea* degree — in between a master's degree and a PhD degree elsewhere — reported average wages of less than E1,500 a month. This is well below what a postdoc in the United States or United Kingdom would expect to earn. It is no surprise that 86% of the respondents want to go outside Italy to acquire new expertise, and that 50% of them think they will have to move abroad to realize their career expectations.

But the young biotechnology community in Italy is getting organized. The ANBI was set up in 2001 to represent young biotech professionals, and has since joined forces with the Young European Biotech Network, which is looking at similar concerns across Europe. ■

Francesco Lescai is president of the Association of Italian Biotechnologists and Marco Quarta is a board member of Young European Biotech Network. Both are at the University of Bologna.

NUTS & BOLTS

A shaken economy and shrinking organizations have created a chilly climate for job seekers. Whether you're a recent graduate, postdoc or career changer, capitalize on your strengths and expedite your search by adopting a system that rewards **knowledge, networks and know-how**. In the inaugural edition of this column devoted to 'nuts and bolts' career strategies for the scientific community, the focus is on **knowledge** and how a combination of self-assessment and trend tracking can help your career.

Self-knowledge is the place to start. Are you best at pure research, or would you rather apply science to practical problem-solving? Would you prefer to focus intensively on one scientific problem or are your interests broader? Do you like working alone or do you value a team environment?

Your answers provide clues to where you're most



With Deb Koen
Careers consultant

likely to thrive. A focus on pure research in a more solitary setting might lead you to academia, whereas solving practical applications in a team is more characteristic of industry. If you're looking for a combination, consider an academic research institute that emphasizes multidisciplinary research or an industry-sponsored institute that focuses on discoveries that could some day be applied. There are no right or wrong answers — simply a case of generating the options that fit your profile.

Once you've identified

the kinds of positions and types of organizations/institutions you're looking for, go beyond advertised openings to scout out opportunities. Every field has its print and online resources. Also use scientific journals, medical directories and business publications to track trends.

Move on to the people. Pick up the phone or arrange a meeting to learn more about a career or an organization of interest. Once you've generated a list of potential employers, start contacting them with the good news of your expertise, enthusiasm and availability. View it as planting the seeds for future opportunities. Even if no opening exists currently, you'll be well positioned when conditions improve.

Deb Koen is vice-president of Career Development Services, a national career management firm, and a columnist for the *Wall Street Journal's CareerJournal.com*. ■

MOVERS

This Week: Mark Boguski, Director, Allen Institute for Brain Science, Seattle, Washington, USA



Some might say that Mark Boguski has a knack for being in the right place at the right time. He landed at the NCBI before the Human Genome Project starting gearing up, and in Seattle just when Microsoft billionaire Paul Allen was planning a new scientific institute. With hindsight those moves seem savvy. But both were risky at the time. No one now doubts the importance of

bioinformatics — whose basic tools Boguski helped to build while at the NCBI. But when he followed his mentor David Lipman to form the NCBI eight months into his postdoc, the term bioinformatics didn't exist, and what little genomic data NCBI generated were distributed quarterly on magnetic tape. Having then moved to Rosetta during the height of human genome fever, Merck's acquisition of the company left him wondering what to do next.

He found it during "what amounted to a sabbatical" at the Fred Hutchinson Cancer Research Center in Seattle, to help them develop a bioinformatics and proteomics strategy. At the NCBI, Boguski realized that bioinformatics by itself was not enough to understand the human genome. At Rosetta, he learned how high-throughput wet biology could combine with bioinformatics to aid drug discovery. And he started wondering how such

industrial technology could serve basic biology. After being introduced to Paul Allen by James Watson, Boguski started consulting on Allen's projected brain-science centre, and last month took the helm when the plans for the \$100-million centre were unveiled (see *Nature* **425**, 226; 2003).

"Going into industry is not the one-way street it used to be."

He imagines some scientists asking: "Why is someone who is not a card-carrying neuroscientist running this project?" He

would answer that all the different skills he picked up along the way — including a genomics mindset and biotechnology management experience from his time at Rosetta — along with a willingness to turn to the institute's scientific advisory board, will help him kick-start the new project. "My story shows that going into industry is not the one-way street it used to be," Boguski says. "You have to be willing to take some risks." ■

CV

2001–2003: Senior director, Vulcan, Seattle, Washington

2000–2001: Senior vice-president, research and development, Rosetta Inpharmatics, Kirkland, Washington

1995–2001: Adjunct professor of molecular biology and genetics at Johns Hopkins University School of Medicine, Baltimore, Maryland

1989–2000: Senior investigator, National Center for Biotechnology Information (NCBI), Bethesda, Maryland

1988–1989: Medical staff fellow, NIDDK, NIH, Bethesda, Maryland